

SQ Sequence 964 AA:
 Query Match 98.2%: Score 4994.5; DB 21; Length 964;
 Best Local Similarity 98.4%: Pred. No. 0;
 Matches 950; Conservative 7; Mismatches 7; Indels 1; Gaps 1;
 QY 1 MKKAFEPFLIGSLGSLAEVPSRITLMPNSVDPPTKESLSKISLTGTHNLNCTYLDN 60
 DB 1 mkkafefllignslsgrlarevpsriltmpnsvdpptkeslskisltdgthnlncylch 60
 QY 61 LRYIIAIILOKTPNEGAAVYITDYLSEFFDQEGYIFAKNLTPESGGAIGYASNSPTVEI 120
 DB 61 lryiiaiioktpnegaaavlitdylseffdqegiyifaknltpesggaiyaspnsptvel 120
 QY 121 RDTICGVIFENNTCCRPFTSSPNNAVKKIRGGALHAONLTIINHHDDYGFKNKFSYR 180
 DB 121 rdticgvifennntccrpfstsspnnaavkkirggalhaonltiinhhddygfknkfsyr 180
 QY 121 rdtlsgpvllennccrcllftwrnpysaa-dkireggaihaqnlyinhhdvvgfknkfsyrg 179
 QY 181 GGAISTANFEVVSSENOSEFLFMDNICIOTNTAGKGAIVAGTSNFSNENCDLFFINMC 240
 DB 180 ggaistanfevvsenoselfmdniciotntagkgaiyagtsnfsnencdlffinnmc 239
 QY 241 CAGGAIFSPICSLTGNRGNIVEYNNRCFKNVETASSEASDGAIKYTRLDVYTGNGRATF 300
 DB 240 caggaiifspicsltgnrgniveymrcfknvelasseasdgaikytrldvyngratf 299
 QY 301 FSDNTITKNGGAIYAPVLTVDNGPTTFINNIANNKGAIVYDGTSSNKSISADRAHAIIRN 360
 DB 300 fsdntitknggaiyapvltvdngptfynnianannkgaiyidgtssnksisadthaiirfn 359
 QY 361 ENIVNTVNTANGSTISANPRRNATITVASSGEIILGAGSSQULIFDYPLEVSNAGVVS 420
 DB 360 enivntvntangstisamprrnatitvaassgeiilgagssquliydyplevsnagvs 419
 QY 421 FKKEADQTSVVFSGATVVSADFHORNLQTKTPAPLTLSNGFLCIDHAQLTVNRFQTQG 480
 DB 420 fkkeadqtsvvsfgatvvsadfhqnlqtktpapltlsngflcidhaqltvnrfqtqg 479
 QY 481 GYVSLGNGAVLSCYKNGACNSASMSSTIKHIGLNLSTLKSQAEIRPLLWEPPTNNNNY 540
 DB 480 gyvslgngavlsckyngtgdssasaeltkhiglnlssllksqaeirpllwepptnnnny 539
 QY 541 TADTATFSLSPVKLSLIDYGNPSPESTDLTHALSSQMLSTSEASDQNRSDDDDFSG 600
 DB 540 tadtaetfslspvklsliddygnspyescdlthalsqpmllstseasdnglqsemdtsg 599
 QY 601 LNVPHYMGGLTWGNAKTQDEPAPASSATITDPKANRFRHTLLTWLPAGVYPSFKHRS 660
 DB 600 lnvphymgltwgwaktdqdepassatitdpqkanrfrhtlltlwlpagvyppskhrs 659
 QY 661 PLIANTLMGNMLATBSLKNASALTPSDHPFMCITGGIGAMVYQDPRENHPCFHRSSG 720
 DB 660 pliantlmgnmlateslknasaeltpsgpfrfwlgtggigamvyqdprenhpgfhrssg 719
 QY 721 YSAGMIAGOTHFSLKFSOTYTKLNERVAKNNVSSKNYSCOGEMLSLOEGFLTLYVG 780
 DB 720 ysagmiagothfslkfsotytklnervaknnvssknyscogemlslsegfltlvlg 779
 QY 781 YSYGDHNCHEFTYTOGENLTSGOTFRSQTMGAVFEDLPKKPEGSTHILTAPELGALGYS 840
 DB 780 ysygdhncchfytogenltsgotfrsqtmgavfედlpkkpegsthilltapelgalgys 839
 QY 841 SLSHFEVATYRSRSTKPLINLVPIGVKGSFMNATORPQAWYELAYQPIYAOBEPG 900
 DB 840 slshfevayprstfckpllnlvpigvkgsfmatbrpqawvelayqpylrbepg 899
 QY 901 IATQULASGIVFGSGSPSRHMASYKISQOTOPLSMLTLHFQYHGFYSSTGCNTLNGE 960
 DB 900 iatqulasgivfgsgspsrhmasysklsqocqplswlthfghfyssstcnylunge 959
 QY 961 IALRF 965
 ILLI

DB 960 ialrf 964
 RESULT 3
 BI3640
 ID BI3640 standard; Protein: 977 AA.
 XX
 AC BI3640:
 XX
 DT 02-FEB-2001 (first entry)
 XX
 DE C. trachomatis pmpe gene amino terminus minus signal sequence protein.
 KW Chlamydial infection; sexually transmitted disease;
 KW pelvic inflammatory disease; PID; tubal obstruction; infertility;
 KW trachoma; blindness; acute respiratory tract infection;
 KW atherosclerosis; coronary heart disease; antibacterial.
 OS Chlamydia trachomatis.
 PN MO200034483-A2.
 XX
 PD 15-JUN-2000.
 XX
 PE 08-DEC-1999; 99MO-US29012.
 PF
 PR 08-DEC-1998; 98US-0208277.
 PR 08-APR-1999; 99US-0288594.
 PR 01-OCT-1999; 99US-0410568.
 PR 22-OCT-1999; 99US-0426571.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;
 XX
 DR WPI: 2000-431303/37.
 XX
 PT Isolated polypeptide for diagnosis and treatment of Chlamydia infection
 PT comprises immunogenic portion of Chlamydia antigen, which comprises
 PT amino acid sequence encoded by polynucleotide sequence .
 XX
 PS Claim 2: Pages 210-212; 256pp; English.
 XX
 CC The present invention relates to new nucleic acid sequences and the
 CC proteins encoded by the nucleic acid sequences. The encoded proteins
 CC comprise an immunogenic portion of a Chlamydia antigen. The encoded
 CC proteins are useful for the serodiagnosis and treatment of Chlamydia
 CC infection. Chlamydiae are intracellular bacterial pathogens that are
 CC responsible for a wide variety of human infections. C. trachomatis
 CC infection is one of the most common sexually transmitted diseases and can
 CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction
 CC and infertility. Trachoma due to ocular infection with C. trachomatis is
 CC the leading cause of preventable blindness worldwide. C. pneumoniae is a
 CC major cause of acute respiratory tract infections in humans and is also
 CC thought to play a role in the pathogenesis of atherosclerosis and
 CC coronary heart disease. The present sequence is a protein isolated in the
 CC present invention.
 CC
 SQ Sequence 977 AA:
 Query Match 96.7%: Score 4919.5; DB 21; Length 977;
 Best Local Similarity 98.4%: Pred. No. 0;
 Matches 935; Conservative 7; Mismatches 7; Indels 1; Gaps 1;
 QY 16 GLAREVPSRIFLMPNSVDPPTKESLSKISLTGTHNLNCTYLDNRYIIAIILOKTPNEG 75
 DB 29 glarevpsriltmpnsvdpptkeslskisltdgthnlncylidnryiiaiioktpneg 88
 QY 76 AAVTTIDVLSFFDQEGYIFAKNLTPESGGAIGVAPNSPVEIRDTIGVPIFENNTCC 135
 DB 89 aavttidvlsffdqegiyifaknltpesggaiyvapnsptvelrtdtignpvlfenncc 148

RESULT 7
ID Y92832 standard; Protein: 1000 AA.
XX Y92832:
AC
XX 07-SEP-2000 (first entry)
DT
XX C. pneumoniae CPN100626 full-length antigen.
DE
XX Antigen: anti-inflammatory; respiratory; antibacterial; anti-asthmatic;
KW anti-arteriosclerotic; vaccine.
XX
XX Chlamydia pneumoniae.
OS
XX MO200024765-A2.
PN
XX 04-MAY-2000.
PD
XX
XX 28-OCT-1999; 99WO-CA000992.
PF
XX 28-OCT-1998; 98US-0106034.
PR 28-OCT-1998; 98US-0106039.
PR 28-OCT-1998; 98US-0106042.
PR 28-OCT-1998; 98US-0106044.
PR 29-OCT-1998; 98US-0106072.
PR 29-OCT-1998; 98US-0106073.
PR 29-OCT-1998; 98US-0106074.
PR 29-OCT-1998; 98US-0106087.
PR 02-NOV-1998; 98US-0106587.
PR 02-NOV-1998; 98US-0106588.
PR 02-NOV-1998; 98US-0107034.
PR 02-NOV-1998; 98US-0107035.
XX
XX (CONN-) CONNAUGHT LAB LTD.
PA
P1 Murdin AD, Oomen RP, Wang J;
XX
XX WPI: 2000-350688/30.
DR N-PSDB: A28710.
XX
XX Chlamydia antigens and the proteins they encode, useful for
PT vaccinating against Chlamydia infections that affect the respiratory
PT tract
XX
XX Claim 13: Fig 21: 226pp: English.
XX
XX The nucleic acids may be used for the recombinant production of the
CC Chlamydia polypeptides (either in vivo or in vitro) according to standard
CC recombinant DNA methodologies. The polypeptides may then be used to
CC vaccinate against Chlamydia infections in mammals. Chlamydia, such as
CC C. pneumoniae, are pathogens responsible for upper respiratory tract
CC infections such as community acquired pneumonia, acute respiratory
CC disease and bronchitis and may be implicated in atherosclerotic changes
CC and asthma. The nucleic acids may also be used as probes for detecting
CC the presence of Chlamydia nucleic acids in samples (and therefore
CC diagnose infections) and the proteins may be used as antigens for the
CC production of antibodies that may be used to detect Chlamydia proteins
CC in samples (e.g. via enzyme linked immunosorbent assay (ELISA)).
XX
XX Sequence 1000 AA:

Query Match 20.8%; Score 1057.5; DB 21: Length 1000;
Best Local Similarity 30.5%; Pred. No. 5,4e-70;
Matches 302; Conservative 166; Mismatches 446; Indels 77; Gaps 28;
OY 6 FFELIGNSLSGLAREVSRITLMPNSVDP-----PTKESLSNK-----ISL 46
AC
XX 56 flllygnftacmfgmcpavyslqdslekfalerdeefrslpdlstltgfsplttf 115
DT
XX 47 TGDTHNLT-NCYLDNLRYLIAILOKTPMEGAAYITTDVLSFFDQKBEITYAKNLTPESG 105
DE

Db 116 vgnrhssqddvlylnyksidnlllwtsgagavscnftl--lsnvedhaflsknlaigt 173
OY 106 GAIGYASPNSPVEIRDTIGPIEFENNTCCRPFTSSNPAAVANKIREGAIHAQ-NLYIN 164
Db 174 gaaacqg----actlktgrplllfssngln-----nastggetrpgalaingdftis 222
OY 165 HNHVDVGFMKNEFSYRGCAISTANTFVVSSENOCELFMDNICIQNTAGKGAIACTSN 224
Db 223 qnqgflfyfvnsvmnvggalstngchrlsqnraplllffn-----ntapsggaltrent 276
OY 225 SPESNNCLFPIFNACCAAGAFISPI-CLSLGNRCNIFYFNWRCRKNVETASSESDGA 283
Db 277 tisdntprlytknccngnggalqtsvtvalknsgsvlfnmtalsg--slngngsgga 334
OY 284 IKVTRLDVGTGNRGRIFESDNTITKNYGAIYAPVYTLVDNGPTPIINNIAANKGAIYID 343
Db 335 l-ytlnjiddnpgyllfnmnyctrdgalcqflltknsghyftm-qgnwggaimgl 392
OY 344 GTSNSKISADRHAILFENNIYTNWNTANGSTSANPPRRMATVASSGCEILLGAGSSON 403
Db 393 qdstclllaegnlafgnnevflltfg-----fynaibhcupns-nlqlganqkyl 441
OY 404 LIFVDPFEVSNAGVS-VSEFNKEADOTGCVSVSGATVNSADFHORMLQRTAPLTLNMGF 462
Db 442 tafldpvehqhtptnplflfnpanhngfllfssaylpeasdyemflsskntselngv 501
OY 463 LCIEDHQAOLTVNRFTOTGTVVSLGAVLSCYKKNAGNSAS-NASITLKHIGLNLSTLK 521
Db 502 lsiedragwqfykftqkqgllkighaaslatansetpsvsgvdlmnaimgl 561
OY 522 SGAEIPILWVEPTNNSNNYADTAATFSLSDVKLSLIDYDGNSEPTSDTLTHALSSQML 581
Db 562 kg-kapliwlrplqssapftednmpdltsq-pllllneenrdpydsldseplqnhll 619
OY 582 SISEASDNOLASDDMDQFGLN-VPHYQWGLMTWCMATODPEPASATITDPOKANRFH 640
Db 620 slsdvlarhlnldhfpeslnatenygyqgiwspwyt--lfttnasi--etanlly 674
OY 641 RTLLTLWLPAGVSPSPKRSPLIANTLWGNMLLATESLKNSEALTPSD--HPFGITGG 698
Db 675 ralyanweplygkvnpeygdlatplwqsfthmftllrsytrgdsdierpfieigia 734
OY 699 LGMWYODPRENHDPGPHNRSSGYS--AGMIAGQHTFSLKFSQTYTKLNERYAKNNVSSK 756
Db 735 dglfvhgnslpgabqfctqyslqasetslqklsldglaqftrtkeisgnvnsah 794
OY 757 N--YSCGEMLFSLQEGFLTKLVGLVSYGDHNCHEFTOGENLTSQTFPSQTMGCAVF 814
Db 795 ntvaslyvelpw-ftgeatatsla-yvgdghlhaylrhkn-raegtcyhtlaaig 851
OY 815 FDLPMKPPGSGNHTLTAPLALGIYSSLSHTEVGAYRSPSTPLNLNVLPVGVGSE 874
Db 852 csfwqgqsyhl--spfvgaiairshqtafeiegndprktvsgpfpnlcipigiqkw 909
OY 875 MNATQRQAWTVELAYOPVLYROEPGIATOLLASKGIFGSGSPSRAMSKYKISQOTOP 934
Db 910 qskfhprewlelsyqpvlyqnpqigvltllaasgswdldghmyvrnalykqhngqal 969
OY 935 LSKWLTMHQYHGFYSSSTFCVYLVNGELALRF 965
Db 970 frsldlfidygsvssstshlqagstlkf 1000

Query Match 20.8%; Score 1057.5; DB 21: Length 1000;
Best Local Similarity 30.5%; Pred. No. 5,4e-70;
Matches 302; Conservative 166; Mismatches 446; Indels 77; Gaps 28;
OY 6 FFELIGNSLSGLAREVSRITLMPNSVDP-----PTKESLSNK-----ISL 46
AC
XX 56 flllygnftacmfgmcpavyslqdslekfalerdeefrslpdlstltgfsplttf 115
DT
XX 47 TGDTHNLT-NCYLDNLRYLIAILOKTPMEGAAYITTDVLSFFDQKBEITYAKNLTPESG 105
DE

RESULT 8
ID Y95551 standard; Protein: 963 AA.
XX Y95551:
AC
XX 10-OCT-2000 (first entry)
DT
XX Chlamydia pneumoniae antigen CPN100624 RY-64.
DE

PD 04-MAY-2000.
 XX 28-OCT-1999; 99WO-CA00992.
 XX 28-OCT-1998; 98US-0106034.
 PR 28-OCT-1998; 98US-0106039.
 PR 28-OCT-1998; 98US-0106042.
 PR 28-OCT-1998; 98US-0106044.
 PR 29-OCT-1998; 98US-0106072.
 PR 29-OCT-1998; 98US-0106073.
 PR 29-OCT-1998; 98US-0106074.
 PR 29-OCT-1998; 98US-0106087.
 PR 02-NOV-1998; 98US-0106587.
 PR 02-NOV-1998; 98US-0106588.
 PR 02-NOV-1998; 98US-0107034.
 PR 02-NOV-1998; 98US-0107035.
 XX (CONN-) CONNAUGHT LAB LTD.
 XX Murdin AD, Oomen RP, Wang J:
 PI WPI: 2000-350688/30.
 DR N-PSDB: A28708, A28709.
 XX Chlamydia antigens and the proteins they encode, useful for
 PT vaccinating against Chlamydia infections that affect the respiratory
 PT tract
 XX Claim 13: Fig 19; 226pp; English.
 PS
 XX The nucleic acids may be used for the recombinant production of the
 CC Chlamydia polypeptides (either in vivo or in vitro) according to standard
 CC recombinant DNA methodologies. The polypeptides may then be used to
 CC vaccinate against Chlamydia infections in mammals. Chlamydia, such as
 CC C. pneumoniae, are pathogens responsible for upper respiratory tract
 CC infections such as community acquired pneumonia, acute respiratory
 CC disease and bronchitis and may be implicated in atherosclerotic changes
 CC and asthma. The nucleic acids may also be used as probes for detecting
 CC the presence of Chlamydia nucleic acids in samples (and therefore
 CC diagnose infections) and the proteins may be used as antigens for the
 CC production of antibodies that may be used to detect Chlamydia proteins
 CC in samples (e.g. via enzyme linked immunosorbant assay (ELISA)).
 CC
 XX Sequence 954 AA:

Query Match 18.6%; Score 947.5; DB 21; Length 954;
 Best Local Similarity 29.1%; Pred. No. 7.8e-62;
 Matches 281; Conservative 166; Mismatches 418; Indels 101; Gaps 34;

OY 35 PKRESLSKISLTGTHNT-----NCYLDNLRLILLOKTPNEGAAVTITDYL 84
 DB 55 plidltln---mtpyshratlfgvrddtngdivldhqnmsiefwefsfsgdgaalscks-1 110
 OY 85 SFEPDOKEGIVFAKNLTPESGAGIYASPNSTVEIRDIGVIEENNCCRPFTSSNPN 144
 DB 111 altnc-kingillinsfalkragam-yvdgn---foisenhgsilltsgnl-----stpn 158
 OY 145 AA--VNKIREGGAIAHONLYINHNHDVYGFMKNFYSVROGAISTNTFTVSENOGCFILPM 202
 DB 159 asnfadctcgvavcskvnltakngtlayflnnkaksaggaalnlhkdltgplcflf- 217
 OY 203 DNICIGTNTAG--KGGAIYAGTSNFSFNCDLFTINNAACCGAGAI--FSPICSLGNRCN 259
 DB 218 -----mnaagtaggaalfanacr-ienmsqpiylfingsgiygallrvhqecllknltgs 270
 OY 260 IYFYNNRCFKNVEFASSEASOGAIVKTRLDVGNRGRIFFSDGNTKNGYATVAPVT 319
 DB 271 vifnmn--fameadfsanhsaggaalycis-csikdnpylaadfmhaardgaalictqst 327
 OY 320 LVDNCPYTFINNIANNKGAIYIDGTSNSKISADRHAIIFENIYTNVNTANGTSTSANP 379
 DB 328 lqdsqpyvftm-qgtwgaalmlrqdgactllfadgdgdliflynnrnfkdlsn--hvsync 384

OY 380 PRRNATVASSGCEILGAGSSONLIFYDPI----EVSNAQSVSPFNKADOTGSVPSG 435
 DB 385 trnvsltyvasgq-----hsatfybdlqrytlqns--lqkfnnpnephllfss 433
 OY 436 -----ATVNSADF--HORNLOKTRPAPLTSNGFCIEBHAOLTVRRFOTGGVSLGAG 488
 DB 434 tyipdstlsrddiffhfrn-----higlyngltaladreaekwykfkqfgyglrlrgsr 486
 OY 489 AVLSCYKNGAGNSASNASITLKHIGLNLSSILKSGAEPLWVPTNNSNNTADTAATF 548
 DB 487 avfstldeeqsssvsgvnlhnlalnlpall-grnvapklrlprgssapyseadnpll 545
 OY 549 SLSDVKLSLIDYGNSPYESTDLTALSSQPMLSSEADNOLRSDMDPSCSLN-VPIHG 607
 DB 546 nlsq-plsllddenldpytdadlaqplaevpillylldvtakhlnthdnfyeglnctqhyg 604
 OY 608 WQGLTWGMAKTPQDEPPASATITDQKANRFRRLTLTLFAGYVPSPKHRSPLIANTL 667
 DB 605 yqgvswpywieclitstdus-----edlvnclhrqlygdwpcptykxvnpenkxdlalsaf 659
 OY 668 W--GNMLATESLKNSEALTPSDHPFWGITGGLGMVYQDPRENHPGFMRRSSGSAG 724
 DB 660 wgsfnlfcaltlyqyqgqlapt-----asgeatrlfwhqsmndakfyhneatqyslg 713
 OY 725 MIA--GQITFSLKFSQYITKLNERAKNNVSKNYSKCGEMLFS-LOEGFLTLKVLGY 781
 DB 714 tsnlasnhsfgvnfsqjfnllyeshsdsvashettvalqjnpwldqerfatsasla-y 772
 OY 782 SYGDNCHHFTYTOGEN--LTSQGTFRSOTMGAVFEDLPMPKFGSTHILTAFLALGY 839
 DB 773 sygn---nhlkasgysgkqteggkystllgaalscsisq-wsrplnfcplqalavr 828
 OY 840 SLSHTEVGAVPRESFRTKPLINLVPIGVGSEFMNATQRPQAVTVELAYQVLYROCP 899
 DB 829 snqatfgeegdkarkfsvbkpnyltplyqisaweskrfplvynlailayqpylyqnp 888
 OY 900 GIATQLLASKGWIFSGSSSSHNAMSYKISQOTOPSLWTLTFQYHNGFYSSSTFCNYLNG 959
 DB 889 einvsiesgsswllsgtlaratalfgynqiflfpklsvldyqgsvsstcthyha 948
 OY 960 EIALRF 965
 DB 949 gtlfkf 954

RESULT 10
 Y92831
 ID Y92831 standard; Protein: 788 AA.

XX Y92831;
 XX 29-AUG-2000 (first entry)
 DE C. pneumoniae CPN100662 processed antigen.
 XX Antigen: anti-inflammatory; respiratory; antibacterial; anti-asthmatic;
 KM anti-arteriosclerotic; vaccine.
 OS Chlamydia pneumoniae.
 PM WO200024765-A2.
 XX 04-MAY-2000.
 XX 28-OCT-1999; 99WO-CA00992.
 PF 28-OCT-1998; 98US-0106034.
 PR 28-OCT-1998; 98US-0106039.
 PR 28-OCT-1998; 98US-0106042.
 PR 28-OCT-1998; 98US-0106044.
 PR 29-OCT-1998; 98US-0106072.
 PR 29-OCT-1998; 98US-0106073.

Query Match 14.3% Score 727; DB 20; Length 822;
 Best Local Similarity 28.1% Pred. No. 1,6645;
 Matches 234; Conservative 140; Mismatches 355; Indels 106; Gaps 35;

35 PTKESLSKISLTGDTNHLT-----NCYLDNRLRYLALLOKTPNEGAATITDVL 84
 48 plidltln---mtyshratlfgyrdndngdividhngmsiesfenfgdgalscks-1 103
 85 SFEDTQKGEIYFAKNLDESGAIGVYASPNSEPTVEIRDITGVIFENNTCCRPFTSSNPN 144
 104 alntc-kngjilfnstakragam-yvngn---fdlsehgslfsgnl-----sfpn 151
 145 AA--VNKIREGGAHNAOLXINHNDVGFPMKNFSYVGAISTANTVVSNSQCFLEM 202
 152 asnfadctggavlesckvltskngtrayflnkkssggaigaalnlkdhltgpltf- 210
 203 DNICLOTWTAG--KGAIVAGTSNFSFESNCDLFFINNACGAGAI-FSPICSLTGNGRN 259
 211 -----maagxtaggaifanacr-iennsqpiyflnngsgagairvghdecltkngs 263
 260 IVFYNNRCFKAVETASSASDGAIKVTRLDVTGNRGRIFFSDNITRNNGAIVAPVVT 319
 264 vlfnn--fameadisahnssggaicys-csikngpjlaafndntaardgaictgalt 320
 320 LVNDGPTFEINNANNKGAIVYIDGTSNKSISADRHAIIFENNTVNTVANGTSTSANP 379
 321 lqdgspvylfn-qglwgaalmrlqdgacclfdagddllfynrhfkdlfn--hvsync 377
 380 PRNRAIVYASSGSEILLGAGSSQNLIFYPDI---EVSNAGVSFVNKEADDTGSVFSG 435
 378 trnsrltvgagq-----hsatfydplllqyltngs--iqfnpnpehlgilits 426
 436 A-----TVNSADF--HQRNLOTKTPAPLTLNSGFLCEDHAQLTVNRPQTGGVSLNG 488
 427 ayldptstrddflshfrn-----higlyngtaldraekwykfdqfgtlrlgsr 479
 489 AVLSCKYKAGNASNASNASTIKHICLNLSTILKSGAELPLMVEPTNNSNNTATATATF 548
 480 avfstldeegsssvgavlnlnalnlpalsl-gnrvapklwlrptgsapysednpll 538
 549 SLSDVKLSLIDDYGNSPYESTDLTHALSSQPMLSISEASDNLREDMDGSLN-VPHYG 607
 539 nlsq-plslldendndpdtadlaqpievpllyldvtakhindntypeglntctqhyg 597
 608 WQGLTWGMAKTODEPARASATITDPOKANRFHRTLLTLWLPAGVSPKHKRSPLIANTL 667
 598 ygvwspwyiaetlitsdts-----edlvnclhrqlygdwprtygkvnpenkgdialasaf 652
 668 W---GNMLLATESLUNKSNALTPSDHFWGIRGGGGMVYDOPRNNHGFHMRSSGYSAG 724
 653 wqstlnlactlrytqgqjapt-----asgeatrlivhqnsmndakgflmneaigyalsg 706
 725 MIA--GQHTFSLKFSQTYTKLINERYAKNNVSKNYSCGEMLS-LOEGFLTLKVLGY 781
 707 tctntashstsfynfsqflnlyeshedsnvashtltvalqinmpwldertstssla-y 765
 782 SYGDHNCHEFTOGEN--LTSQGFTRSQTM-GGAVFPLPKKPGSTHILAP 831
 766 sysn---hthkasgyssgiktegcystllrgsllsls-----stmaaltcp 810

RESULT 12
 Y16738
 ID Y16738 standard; Protein: 1013 AA.
 AC Y16738;
 XX
 XX 21-JUL-1999 (first entry)
 DE C. trachomatis F serovar HMW protein.
 XX
 XX Chlamydia: high molecular weight protein; HMW protein; urethritis;

KW bacterial infection; conjunctivitis; lymphogranuloma venereum; LGV;
 KW cervicitis; epididymitis; endometritis; pelvic inflammatory disease;
 KW PID; salpingitis; tubal occlusion; infertility; cervical cancer;
 KW arteriosclerosis; atherosclerosis.
 OS Chlamydia trachomatis.
 XX
 XX W09917741-A1.
 PN
 XX 15-APR-1999.
 PD
 XX 01-OCT-1998; 98WO-US20737.
 PF
 XX 02-OCT-1997; 97US-0942596.
 PR
 XX (ANTE-) ANTEX BIOLOGICS INC.
 PA
 XX Jackson JW, Pace JL;
 PI WPI: 1999-287659/24.
 DR
 XX
 XX
 XX
 XX
 PS Claim 4; Page 119-123; 141pp; English.

The invention relates to an isolated Chlamydia species high molecular weight (HMW) protein having an apparent mol. wt. of 105-115 kD as determined by SDS-PAGE. The HMW proteins and nucleic acid molecules can be used for preventing, treating or ameliorating a disorder related to Chlamydia e.g. bacterial infection, conjunctivitis, urethritis, lymphogranuloma venereum (LGV), cervicitis, epididymitis, endometritis, pelvic inflammatory disease (PID), salpingitis, tubal occlusion, cervical cancer, infertility, arteriosclerosis and atherosclerosis. The products can also be used for detection and diagnosis. The present sequence represents a C. trachomatis HMW protein.

Sequence 1013 AA;

Query Match 12.8% Score 651.5; DB 20; Length 1013;
 Best Local Similarity 25.6% Pred. No. 9,2640;
 Matches 260; Conservative 152; Mismatches 389; Indels 213; Gaps 43;

39 SLNKSISLTGDTNHLNCLNRLRYLALLOKTPNEGAATITDLSFEDTQKEIYFAK 98
 84 nlqglfvlgrghslt---fenir-----tsungals-----dsanglf--- 121
 99 NLPESGAIGVYASPNSEPTVEIRDITGVIFENNTCCRPFTSSNP-NAAV----- 147
 122 -tleqfelstfscnslavl---paatlmgsgqlptltstpsngltlyskldl111n 174
 148 -----NKIR-EGGAHNAOLXINHNDVGFPMKNFSYVGAISTANTVVSNSQCFLEM 198
 175 nekstfynlvsogdgtldaksilcvqglsklcvfgentadagdgagcyvtsfsamnap 234
 199 FLEMDNICIQTNTAG-KGAIVA-----GTSNSEF-----SNDCLEFFINNACGAC 243
 225 laflanv-----agvryggaavdgqgvsststedprvstfntlaveldgvarvg 288
 244 GAIFPICSILTGNRCNIVFYNN--RCFKN-----VETASSASDGAIGAI 284
 289 991ys-----ygnvafllngkclflnvaaspyiaaegptngqaantsdnyddga1 340
 285 KVTTRLDVTGN-----RGRIFFSDNITRNNGAIVAPVTVLVNCPYFFINNTAN 334
 341 fckngagaagsnsgsvsfddgeyvvfssnvaagkggaikylklsivancgpyvflgnlan 400
 335 NKGAIVYIDGTSNKSISADRHAIIFENNI---VTNVTNANGTSTSANPPRRAITVASS 390.
 401 d-ggailygsegeisltsadylldgnlkrtakenaadvngvtvss-----qalsmgsg 454
 391 SGELLGAGSSQNLIFYPFIEVSN-----AGVSVSFNKEADDTGSVFGAIVNSADF 443

693 GTTGGGLMAYVQDPRNPRGFNMSSGYS--AGMAGOTHTFSLKFSQTYTKLNERPAK 750
OY
Db 641 -----giadlfndeqngnrsyrhsasgylggfftaesnfafccqlf-----gydk 690
OY 751 NNVSSKNS--CGEMLF-SLOEGFLTKLVG-----LYSGDHNCNHFT 793
Db 691 dhlvaahnhvuyagamsyrlhgesktlaklsgnsdrlfvinarlayutndmcttkyt 750
OY 794 OCENLTSQGTFRSQTM---GGAVFEDLPMKPRFG---STHLLAPFLGALGIYSSLH 844
Db 751 gysprv--kxswndatgiccgai-----pvasgrtswdth---cpfinlemiyahqnd 801
OY 845 FTEVGAYPRFSSTKTPRLVLPVIGVSGFMNATORPQAMVELAYOPVLYROEPIATQ 904
Db 802 fkeugfegrsfgse-dlfnlavpvgik--fekfsdk-stydlslayvpvlnrdpccitl 857
OY 905 LASKQIMFGSGSPSRHAMSRYKISQOTRPLSMVLTHFQYHGFSSSTCNLYNGSLAR 964
Db 858 lmvsgdswstcglstsrqalvlvrag-----nhaf--asnfevisqfevelr 902

RESULT 14
B13633 standard; Protein: 982 AA.
B13633:
02-FEB-2001 (first entry)
C. trachomatis pmpG gene protein.
Chlamydia infection; sexually transmitted disease;
pelvic inflammatory disease; PID; tubal obstruction; infertility;
trachoma; blindness; acute respiratory tract infection;
atherosclerosis; coronary heart disease; antibacterial.
Chlamydia trachomatis.
Key Location/Qualifiers
MISC-difference 981 /note="Unspecified amino acid"

08-DEC-1999; 99WO-US29012.
08-DEC-1998; 9805-0208277.
08-APR-1999; 9905-0288594.
01-OCT-1999; 9905-0410568.
22-OCT-1999; 9905-0426571.
(COR1-) CORIXA CORP.
Probst P, Bhatia A, Skeiky YAM, Fling SP, Jen S, Stromberg EJ;
WPI: 2000-431303/37.
Isolated polypeptide for diagnosis and treatment of Chlamydia infection
comprises immunogenic portion of Chlamydia antigen, which comprises
amino acid sequence encoded by polynucleotide sequence -
Claim 2: Pages 181-184; 256pp; English.
The present invention relates to new nucleic acid sequences and the
proteins encoded by the nucleic acid sequences. The encoded proteins
comprise an immunogenic portion of a Chlamydia antigen. The encoded
proteins are useful for the serodiagnosis and treatment of Chlamydia
infection. Chlamydiae are intracellular bacterial pathogens that are
responsible for a wide variety of human infections. C. trachomatis
infection is one of the most common sexually transmitted diseases and can

lead to pelvic inflammatory disease (PID), resulting in tubal obstruction
and infertility. Trachoma due to ocular infection with C. trachomatis is
the leading cause of preventable blindness worldwide. C. pneumonia is a
major cause of acute respiratory tract infections in humans and is also
thought to play a role in the pathogenesis of atherosclerosis and
coronary heart disease. The present sequence is a protein isolated in the
present invention.
Sequence 982 AA:
SQ
Query Match 12.7%; Score 647; DB 21; Length 982;
Best Local Similarity 25.4%; Pred. No. 1,9e-39;
Matches 257; Conservative 156; Mismatches 388; Indels 212; Gaps 43;
39 SLSNKSLTGDTHNLNCLDNLRLYALLOKTPNGAAVTIRDYLSFDTOKEGIYPAK 98
Db 54 nlgsftvigrhslc--fenlr-----stngaals-----nsaagdlf--- 91
OY 99 NLPESGAIGYASPNSPVEIRDTIGPVIENNTRCPFTSSNP-NAAV----- 147
Db 92 --tlegfkelstfencnslavl-----paatnkgsqrpdtletspngtlysktdllln 144
OY 148 -----NKIR-EGGAIHAONLXINHNHDYGFPMKNFSYRGCALSTANTPVVSNOSC 198
Db 145 nekfsfysnlvsgdggaidskltvgjsklvfgentagdgagcqvvtlsamaneap 204
OY 199 FLEMDNICIQTNAG-KGGAIVA-----GTSNSFE-----SNNCDFEINNACGAG 243
Db 205 iaftvanv-----agvrggjaavdgqgqvssstsedpvsftrntavefgnvarvg 258
OY 244 GAIPLPISGLTGNRGNIVFVNN--RCFKN-----VETASSASDGA 284
Db 259 gqlys-----ygnvatfmgkklflnvasepylaakqptsgqasnlsmnygggal 310
OY 285 KVTTRLDVYGN-----RCRIFESDNTKMYGGAIVAPVTVLNDGPTVYINNANN 335
Db 311 fcknqagagsgnsvsfqdgvgvffsnaagkgajakklsvanogpvgqlrnlano 370
OY 336 KGGAIVIDGTSNKSISADRIHAIFFENI-----VTNNTNNGSTSTANPPRNAITVASSS 391
Db 371 -ggaitylgesgelsiaadgdlifdgnlkrakenaadnngvrvss-----qaisngsgg 424
OY 392 GELLIGAGSSQNLIFYDPIEVSN-----AGVSVPKNEADOTGSGVFSGATVNSADPH 444
Db 425 klttlrakaghqllfndplemangnqpqgskllikndegyltgdlvfan--gssltly 481
OY 445 QRNLQTKTAPRLTSLNGFLICIEDHAOLTVKRFTOTGCVSLGCAVLSCYKKNAGS--A 502
Db 482 qn-----vlieqgrivltekaklsvnsisqtsqslymeegstldfvtppqpqpa 532
OY 503 SNASITLKHIGLNLSSILSGAEIPLMWEPN--NSNNTATATATFSLDVKLS--L 557
Db 533 anqjlitlsmhslssellanna-----vtrpptnppagdsipavigt--tagsttisgplf 587
OY 558 IDDYGNSPYESTD-----LTHALSQPMLSISEASDNQLRSDMDQFGLNVPHYG 607
Db 588 fedldtdvdyrwlgnsqknvklqigtkp--panapadlcl-----gnempkyg 637
OY 608 WQGLMTWCAKATODPEPASATITDPOKANRFRHTLLTLVLPAGVYSPKRRSPILANTL 667
Db 638 yggswklaw-----dptannprrlkatvktgynpppervaslvpnsl 682
OY 668 WGNMLLATESLKNSADELPDSHDFM--GITGGGLMAYVQDPRNPRGFNMSSGYSAGMI 726
Db 683 wgs-ildirsahaeiqasvdgrycrglwwsgvsnfftyndrdaiggyuyisgyslg-- 739
OY 727 AGOTHTFSLKFSQTYTKLNERPAKNNV--SSKNVSCGEMLFSLQECFLTKLVGLYSTG 784
Db 740 -ansygsnmgfajaftevfgr-skdyvvrnsnhacisgvyaltqga----lcsyylfig 792
OY 785 DHNCHHFTYOG-ENLTSQGTFRSOT-----MGCAVFFEDLPMKPRGSHILT--APFL 833

Db 678 dfeenifsvafcgldfgkdkdlfiven-tshnylaelylqhraflgglpmpsfsgsltdmlk 736

Oy 770 EGFLLTKLVGYSYGDHNCHEFT-----OGENLTSGQTFRSQTMGAVFFDLPMK-PFG 823

Db 737 dipllnaqlsystkndmdrlytsypaagswlnsgale---lgsalalylkpeaplf 793

Oy 824 STHTLAPFLGALGYSSLSHFTVEGAYPRSFSTKPTPLINLVLPICVKGSPFNATQROQA 883

Db 794 qgyf---pflkfqavysrqnfkesgaearafd-dgdlvncslpvglrleklsedek-nn 848

Oy 884 WTVELAQPVLYRQEPGATQOLLSKGIW 912

Db 849 feislaiygdvyrknprstslmvsasw 877

RESULT 23

W88424

ID W88424 standard; Protein; 930 AA.

AC W88424;

XX 26-APR-1999 (first entry)

DT

XX Chlamydia pneumoniae surface exposed protein Omp11.

DE

XX Omp11: outer membrane protein 11; surface exposed protein;

KM antigen; infection; diagnosis; vaccine; atherosclerosis; asthma.

XX Chlamydia pneumoniae.

OS

XX MO9858953-A2.

PN

XX 30-DEC-1998.

PD

XX 19-JUN-1998: 98WO-DK00266.

PF

XX 23-JUN-1997: 97DK-0000744.

PR

XX (BIRK/) BIRKELUND S.

PA (CHRL/) CHRISTIANSEN G.

PI Birkelund S, Christiansen G, Knudsen K, Madsen A;

PI Myding P;

XX WPI: 1999-105610/09.

DR N-PSDB: X06823.

XX

XX Species-specific test for identifying mammals infected with

PT Chlamydia pneumoniae - comprises detecting antibodies specific for

PT outer membrane proteins of C. pneumoniae or nucleic acids encoding

PT these proteins

XX

PS Claim 7; Page 63-65; 115pp: English.

XX

XX This polypeptide comprises the novel 97.6 kDa surface exposed

CC protein Omp11 of the human respiratory pathogen Chlamydia

CC pneumoniae. Its amino acid sequence was deduced from DNA (see

CC X06823). Isolated from a C. pneumoniae expression library. The

CC invention provides 12 novel surface exposed proteins. Omp4-Omp15

CC (see W88417-28), and nucleic acid sequences encoding them (see

CC X06816-27). A new species specific test is claimed that is used

CC to identify mammals (including humans) infected with Chlamydia

CC pneumoniae. The test comprises detecting antibodies specific for

CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer

CC membrane proteins, especially by PCR. The proteins are also used

CC in the diagnosis of C. pneumoniae infection in mammals. The

CC nucleic acids and proteins can also be used in the immunization of

CC mammals, the nucleic acids being particularly useful as DNA

CC vaccines for effecting in vivo expression of antigens. The

CC vaccines may also prevent atherosclerosis and bronchial asthma,

CC which are possibly associated with C. pneumoniae.

XX

XX Sequence 930 AA;

XX

XX

Query Match 12.0%; Score 608; DB 20; Length 930;

Best Local Similarity 26.9%; Pred. No. 1.4e-36;

Matches 250; Conservative 140; Mismatches 375; Indels 164; Gaps 40;

Oy 47 TGDTHNLNLCVYIDNRLRYLALIQKTPNEGAAVT---INDYLSFTQKQEGIFYAKNLTP 103

Db 50 tada-nglnvylsgnvyi-----ndagxqtlbgcftletgdlffgkyfsfsh-lvd 102

Oy 104 SCGAIGVASPNSPIVEIRDT---IGPVIFFENNTCCRPFTSSNPAAANKIREGGAHAQ 159

Db 103 agsmagaasttadkalfctgfnsfifaagptvaasgkstssagalnlnclngclifsg 162

Oy 160 NLYINHHIDVGFMKNFYVKGAISTANTFVSENQSCFLFMDNICIQTNTAGK-GGAI 218

Db 163 nvsneann-----ggaflt-kltslsgnts-----slftsnakklggal 203

Oy 219 YAGTNSFESNCDLFTINN-ACCAGGAI-FSPICSLGNGNIVFYNNRCFKNVETASS 276

Db 204 yssaaaslsngtlqglvlnmkgetlgygalgteassslqgn-sllfsgn-----lald 255

Oy 277 EASDGAIKV-----TTRLDVATGNRGRIFFSDNIRKNGAIVAVVTLVNDGPTPIFN 331

Db 256 aagkgalcyckctgetpctltisgnks-ltfaensvtggaicahglolsaagpctlsfm 314

Oy 332 IANN---KGAIIYIDGTSNSKISADRHAIIFENENIVNTNANGSTSTANPPRRNATV 387

Db 315 rcgnlaagkggaladsgslsangsdlcflgnrl-----tstscptlstrnaly 366

Oy 388 ASSSGEILLGAGSSONLIFYDPIEVSNAQVS-----VSFNKEADQGSVYFSGATVNS 440

Db 367 gssaklalnraqgslyfydpilasnlqgsadvllingpdsnpldyqglvifsgkls-s 425

Oy 441 ADFIQR--NLQTKTPAPVTLSENGFLCIEDHMQLTYNRFOTQGVVSLNGAVLSQYKKA 498

Db 426 adeakaadnftcsllkqplalasgtlalkgnveldvngltclegstll-----mqgyl 477

Oy 499 GNSASNAITLKHIGLINSITLKSAGAEIPLLWVEPTNNSNNYADTAFTSLSDVKSLI 558

Db 478 klkadleaalsllkvlvdlsal-----egnkavsietlaganklllts-plvfq 524

Oy 559 DDYGNPVESTDLTHALSSQPMLSISEASDQNLSDQMDPFGCLAV-----PHYGQGL 611

Db 525 dssgnf-yeshntlnqat-rcpdlvflaat-----aasdiyalallspvqrpephyygsh 578

Oy 612 WTGWAKTQDPEPASATITPOKANRFRHTLLTWLPAGVPSPKHRSPLIANTLQNM 671

Db 579 weatwa---dstaksgt-----mtwvltgynpmperrtasvpslwasf 620

Oy 672 LLATESLKNASAEITPSDHPFV---GITGGGLGMVYODPRENHPGFHRRSSGYSAGMIAG 728

Db 621 ---cdirltqlqmtsqaanslyqgqlwasgylanffhkdksglnqaftrkxsgylvgssae 677

Oy 729 --QHTPSLAKRSQVYTKINERYAKNNVSKNYSCO-----GEMLSLQ 769

Db 678 dfeenifsvafcgldfgkdkdlfiven-tshnylaelylqhraflgglpmpsfsgsltdmlk 736

Oy 770 EGFLLTKLVGYSYGDHNCHEFT-----OGENLTSGQTFRSQTMGAVFFDLPMK-PFG 823

Db 737 dipllnaqlsystkndmdrlytsypaagswlnsgale---lgsalalylkpeaplf 793

Oy 824 STHTLAPFLGALGYSSLSHFTVEGAYPRSFSTKPTPLINLVLPICVKGSPFNATQROQA 883

Db 794 qgyf---pflkfqavysrqnfkesgaearafd-dgdlvncslpvglrleklsedek-nn 848

Oy 884 WTVELAQPVLYRQEPGATQOLLSKGIW 912

Db 849 feislaiygdvyrknprstslmvsasw 877

RESULT 24

W88417

XX WO200032794-A2.
XX
XX 08-JUN-2000.
PD
XX 01-DEC-1999: 99W0-CA01147.
PF
XX 01-DEC-1998: 98US-0110339.
PR
XX 01-DEC-1998: 98US-0110340.
PR
XX 01-DEC-1998: 98US-0110427.
PR
XX 01-DEC-1998: 98US-0110428.
PR
XX 01-DEC-1998: 98US-0110438.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Murdin AD, Oomen RP, Wang J;
XX
XX WPT: 2000-412339/35.
DR
XX N-PSDB: A30847, A30848.
DR
XX
XX Nucleic acids encoding polypeptide antigens from Chlamydia useful for
PT preventing, diagnosing and treating diseases such as community acquired
PT pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset
PT asthma -
XX
XX Claim 16; Fig 1; 174pp; English.
PS
XX
XX This sequence is a Chlamydia antigen of the invention, designated
CC CPN100634. The nucleic acids (and their complementary sequences) may be
CC used as diagnostic agents for detecting the presence of nucleic acids
CC encoding Chlamydia antigens in samples according to standard methods,
CC and therefore, for diagnosing Chlamydia infections. For example, they may
CC be used as primers and probes for diagnostic polymerase chain reaction
CC (PCR) assays. Antisense sequences may be used to down regulate
CC expression of the proteins and may be used to treat infections. The
CC nucleic acids may also be used to produce the protein antigens they
CC encode according to standard recombinant DNA methodologies. The
CC proteins may then be used as antigens for the production of antibodies
CC (i.e. as vaccines) for preventing infection by Chlamydia. The
CC antibodies may also be used as diagnostic reagents for detecting
CC infections. Chlamydia is a pathogen implicated in the development of
CC (for example) community acquired pneumonia, upper respiratory tract
CC disease (especially bronchitis and sinusitis, asthmatic bronchitis,
CC adult-onset asthma and acute exacerbations of asthma in adults.
XX
XX Sequence 928 AA:
SQ

Query Match 11.7%; Score 597; DB 21; Length 928;

Best Local Similarity 26.1%; Pred. No. 9, 2e-36;
Matches 253; Conservative 148; Mismatches 361; Indels 206; Gaps 47;

QY 36 TKESLSNKLSTGTD-----TNLITNCY--LDNLRVY-----LAILQKTPNEGA 76
DB 45 lpktsatysltgdyfeyepgktpldscfktcdnlftlgnhslftgfdaglhaga 104
QY 77 AVTIDVLSFFDTCKEGITFAKNLTPESGCAIGVASPSPTVEIRDITGCVLFENNCTCR 136
DB 105 aastc-----anklftsgfsllstfsspssttv-----ltgq----- 136
QY 137 PFTSSNPAAVKKIREGAIHQNLYNNHNDVGFMMKFSYVRGCAISTANTFVASENQ 196
DB 137 -----gltsaagvnlentl---rklyvag---nftsaadgaikga-sllltgts 178
QY 197 SCFLPMDNICIOTNTAGAGAIYAGTSNFSFENNCDL--FTINNACGAGAIISPICSLT 254
DB 179 gdalifsm-----ssstkgal-attagariantgyvrflianlastsgalddegtsll 232
QY 255 CNRGNIVFYNNRCFNVETASSEASDCAIKVTRRID---VTGNRGRIFFSDNTKXNG 310
DB 233 snnkfllyfegn-----aaktggaicntkassgsPELLISnktllifasnvaetsg 282
QY 311 GAIVAPVTVLDNCGPTFTI-NNI--ANNKGAIYIDGTSNSKISADRHAIIFENENIVTVN 367

DB 283 gahhakkllalsggyfcellinnvssatpkygalsidaesgelsisaetgnltfvrnlt-- 340
QY 368 TNANGTSTANPPRRNAITVAVSSGCEILLAGSSQNIlfyDPI-EVSNAGVSVSFNKEA 425
DB 341 -----ltgstcdpkrnalnlgnsngkftelraaknhliffydphtsestsadvlikingaa 395
QY 426 ----DOTGSVTFSGATVNSADFH-QRNLOTKTPAPLITLSNGFICIDHAOLYVNRTO-T 479
DB 396 galnpyqgtlltsgecltadelkvadnlkssftqpvslsgklllqkyvtltestsfsgaa 455
QY 480 GGVVSLGNGAVLSCYKNGAGNSASNSITLKNHGLUSSILKSGATPIPLMWEPTNSNN 539
DB 456 galldmgsclt-----stagslltlnlglnvdl---glkqv-----s 493
QY 540 YTAADTAATFSLSDVKLSLIDYGNSPYESTDLTHALSSQPMLSISEASQNLRSDDMFS 599
DB 494 ltagasnkvivsgklnlidlagnl-yes----hmftshdqllslklvldadvtlnvds 548
QY 600 GL-NVP-----HYGQGLWTGMAKTQDPERASSATITDPOKANFHTLLTLWLPAGY 652
DB 549 slpvpaeqpnseygftqgwnvnm-----tltalatnkteat-----atwktgft 592
QY 653 VPSPKHSPLINLTWGNMLLATESLKNSAEL--TPSDHP--FWGITGGGLGMVYQDPR 708
DB 593 vpspeksalvncllwg-vftdtrslqqlveigatcmehkgfiv-----vssmtnflnklgq 648
QY 709 EHNHPCFHMRRSGVSAQMIAGOTHT-----FSLKFSQTYTKLMERY-AKNN----- 752
DB 649 entrgftrhsygy----vlgssahtrpkddlftafchlfardcdcfiahnnstlygltff 705
QY 753 -----VSSKNYSCQGEMLFSLQ--EGF-----LTKLVLYSYGDHNCHEFTYQGENLTS 800
DB 706 khshtlqpnylirlgrakfsesaiekfpreipjaldivgsfshsdmrmethyslpe--s 763
QY 801 OGCFRSGTGWGAVFDPDEPKPFGSTHILAPPLGALG---IYSSLHFTFEVAGAPRSFST 857
DB 764 egswsneciaaglgldlpt-vlsnphlilkltlppmkvemvvsqnsfsssdggrfsl 822
QY 858 KTPILNVLPICVKGSEFNMATQRPQAMTVELA--YQPVLYROEPCGIATQLLASKGIWPCS 915
DB 823 gr-llnlsipgak--flvg-dlgdsytydlsgffvysdvyrnmpgstallvmspdswwlr 878
QY 916 GSPSSRHA 923
DB 879 ggnlsrqa 886

Search completed: May 6, 2001, 19:16:43
Job time: 4308 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 19:17:30 ; Search time 59.76 Seconds

(Without alignments)
553.155 Million cell updates/sec

Title: US-09-677-752-2

Perfect score: 5086
Sequence: 1 MKKAFPFLLIGNSLSGLARE.....GVSSSTPCNYLNGELALRF 965

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4989.5	98.1	964	1	OMPE_CHLTR
2	866.5	17.0	1034	1	OMPE_CHLTR
3	220.5	4.3	1250	1	YFAL_ECOLI
4	193	3.8	1643	1	OMPB_RICPR
5	191.5	3.8	1645	1	OMPB_RICPR
6	190	3.7	1300	1	120K_RICRI
7	190	3.7	1654	1	OMPB_RICRI
8	189	3.7	1656	1	OMPB_RICJA
9	178	3.5	2003	1	YDBA_ECOLI
10	174	3.4	2249	1	190K_RICRI
11	173	3.4	1310	1	VAC3_HELPY
12	170.5	3.4	2329	1	YS89_GABEL
13	161	3.2	1286	1	AIDA_ECOLI
14	159.5	3.1	1385	1	P3K2_DICDI
15	157.5	3.1	1325	1	YDER_ECOLI
16	155	3.0	1569	1	YPUA_ECOLI
17	153.5	3.0	593	1	CSG_METFE
18	153.5	3.0	593	1	CSG_METFE
19	150	2.9	976	1	F1PB_ADEB3
20	147	2.9	1039	1	AG43_ECOLI
21	146.5	2.9	1025	1	SLAP_CAUCR
22	145.5	2.9	863	1	YEOJ_ECOLI
23	143.5	2.8	995	1	Y109_YEAST
24	141.5	2.8	959	1	N100_YEAST
25	140	2.8	1117	1	YNG6_YEAST
26	139	2.7	1006	1	BGAL_ASPNG
27	139	2.7	1041	1	EGT2_YEAST
28	139	2.7	1394	1	HAP_HAETN
29	139	2.7	1419	1	ALAI_CANAL
30	138.5	2.7	1169	1	YK82_YEAST
31	138	2.7	1004	1	SLPO_BACBR
32	138	2.7	1026	1	VG37_BPT4
33	138	2.7	1902	1	P2P_LACPA

34	137.5	2.7	1243	1	VG37_BPK3	O38394 bacterioph
35	137.5	2.7	1306	1	MSB2_YEAST	P32334 saccharomyc
36	137	2.7	1577	1	HLVA_PROMI	P16466 proteus mir
37	137	2.7	1902	1	P1P_LACLC	P16271 lactococcus
38	136	2.7	1288	1	VACA_HELPD	O924W5 helicobacte
39	135.5	2.7	734	1	G13B_DICDI	P34116 dictyostell
40	135.5	2.7	2278	1	FAB1_YEAST	P34756 saccharomyc
41	134.5	2.6	827	1	XAMP_XANS2	O60106 xanthomonas
42	133	2.6	704	1	OE66_NPVAC	O00704 autographa
43	133	2.6	881	1	YJH8_YEAST	P47033 saccharomyc
44	132.5	2.6	580	1	FLA2_PYRKO	O9V2X0 pyrococcus
45	131.5	2.6	928	1	HXA2_HAETN	P45354 haemophilus

ALIGNMENTS

RESULT	ID	OMPE_CHLTR	STANDARD	PRT	964 AA.
AC	O84877:				
DT	30-MAY-2000 (Rel. 39, Created)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	POTATIVE OUTER MEMBRANE PROTEIN F. PRECURSOR.				
GN	PMPE OR CT869.				
OS	Chlamydia trachomatis.				
OC	Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.				
OX	NCBI_TaxID=813:				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=D/UV-3/CX;				
RC	MEDLINE=99000809; PubMed=9784136;				
RA	Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Atavind L.,				
RA	Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,				
RA	Davis R.W.:				
RT	"Genome sequence of an obligate intracellular pathogen of humans:				
RT	Chlamydia trachomatis."				
RL	Science 282:754-759(1998).				
CC	-1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)				
CC	(POTENTIAL).				
CC					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; AED01360; AAC68467.1; ..				
KM	Outer membrane; Signal.				
FT	SIGNAL 1 18				
FT	CHAIN 19 964				
SO	SEQUENCE 964 AA: 104703 MW: 1899847D2571CE2 CRC64:				

Query Match Best Local Similarity 98.1%; Score 4989.5; DB 1; Length 964;
Matches 949; Conservative 7; Mismatches 8; Indels 1; Gaps 1;

QY	1	MKKAFPFLLIGNSLSGLAREVPSRIPLMPNSVDPPTRESISNKLSTLGDTHNLNCLDN	60
DB	1	MKKAFPFLLIGNSLSGLAREVPSRIPLMPNSVDPPTRESISNKLSTLGDTHNLNCLDN	60
QY	61	LRYLAILOKTPNGCAVTTIDYLSFDTQEGIFPKNLTPESGAIGVAPSPVTEI	120
DB	61	LRYLAILOKTPNGCAVTTIDYLSFDTQEGIFPKNLTPESGAIGVAPSPVTEI	120
QY	121	RDITGPIVENNTCCRPFTSNPNAANKIREGCAIHAONLYINHNDVYGFKNFSYVR	180
DB	121	RDITGPIVENNTCCRPFTSNPNAANKIREGCAIHAONLYINHNDVYGFKNFSYVR	180

```

Db 555 NPODMONTGTGTGTPSSSTSSISTPMIIFNGRLSIYDENVESYDMSDISRGKADOLLI 614
Oy 584 SEASDNOLRSO---DMDFSGNLVPHYGMGLMTWGMAKT-----ODPPASSATTTD 632
Db 615 ETTDGDOLDSSWSSSLNLSLSPHYGLOGLMTPTMTTITTLNNSSAPTSATSTAE 674
Oy 633 POKANR-----FHRLLLTWLPAGVYPSFKH 658
Db 675 QKKTSETFTPSNTTASIPNIKASAGSGSASNGEVTITKHTLVVMAPGVYDPIR 734
Oy 659 RSLPIANTL---WGMMLLATESLKASALTPSDHPFNGITGGGLGMVYODPREHHPFH 715
Db 735 RGDLLANSLVHSGRMWTGLRSL-----PDNSWFALOGAATTLFTKQORRLSYHSTIS 787
Oy 716 MRSSYSAGMIA--GQTFEFLKFSQTYTKLINEYAKNNVSSKNY---SCGEMFLSQE 770
Db 788 SASKGYTVSSQASGAHGHKFLSFSOSSDKMEKETNNRLSRYYLSALCFHPWFD--- 844
Oy 771 GFLTRKLVG--LYSYGDHNCHEFTGTGENTLSQGTFRSQTMGAFFDL---PMKPFQS 824
Db 845 ---RALIGAAACNGTGNHMRSF--GTKKSSKGKHFSTTGASLRCELDSMPLR--- 895
Oy 825 THILTAPELGALGIYSLSHFTFEGVAYPRSGSTKTPILNVLPVIGVKGFSMATQRPAW 884
Db 896 -SIMLTPFOALFSTREPASIRESGDLARLFTLEQAHTRAVSPISIKGAYSSDTPTLISW 954
Oy 885 TWELAYOPVLYNOEGDIATOLLASGIMFGSPSSPSRMSYKISQOTPLMLTLHFQY 944
Db 955 EMBL:Y00544: BAA16052.1: ALT_INIT.
Oy 945 HGFYSSTFCNLYNCEIALRF 965
Db 1014 QAEVATSTVSHYINNAGCALVF 1034

RESULT 3
YFAL_ECOLI STANDARD: PRT: 1250 AA.
AC P45508: P45507: P45506: P39441: P76468: P77487:
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 131.2 KDA PROTEIN IN UBIS-NRDA INTERGENIC REGION
DE PRECURSOR.
GN YFAL.
OS Escherichia coli.
OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
OC Escherichia.
OX NCBI_TaxID=562:
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655:
RC MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12:
RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiuchi T.,
RA Ikeno K., Inada T., Isono K., Isono S., Itoh T., Kanai K., Kasai H.,
RA Kashimoto K., Kim S., Kimura S., Kitagawa M., Kitakawa M., Makino K.,
RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
RA Tagami H., Takekoshi K., Wada C., Yamamoto Y., Yano M.;
RA Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP PRELIMINARY SEQUENCE OF 1-938 FROM N.A.
RX MEDLINE-84272624; PubMed-6087316;
RA Carlson J., Fuchs J.A., Messing J.;

```

```

RT "Primary structure of the Escherichia coli ribonucleoside diphosphate
RT reductase operon."
RT Proc. Natl. Acad. Sci. U.S.A. 81:4294-4297(1984).
RN [4]
RP SEQUENCE OF 925-1198 FROM N.A.
RC STRAIN-K12 / EMG2:
RA Escier P., O'Keefe T., Robison K., Church G.M.;
RN Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1180-1250 FROM N.A.
RC STRAIN-0V6:
RX MEDLINE-88201664; PubMed-2834621;
RA Hussain K., Elliott E.J., Salmond G.P.C.;
RT "The parD-mutant of Escherichia coli also carries a gyrA mutation.
RT Mol. Microbiol. 1:259-273(1987).
RN [6]
RP IDENTIFICATION.
RX MEDLINE-96032851; PubMed-7567469;
RA Borodovsky M., McIninch J., Koonin E.V., Rudd K.E., Medigue C.,
RA Danchin A.;
RT "Detection of new genes in a bacterial genome using Markov models for
RT three gene classes."
RL Nucleic Acids Res. 23:3554-3562(1995).
CC CC -I- SIMILARITY: TO E.COLI YDEK.
CC CC -I- CAUTION: REF.3 SEQUENCE DIFFERS EXTENSIVELY FROM THAT SHOWN BY
CC CC MANY FRAMESHIFTS.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL: AE000313: AAC75293.1:
DR EMBL: D90855: BAA16052.1: ALT_INIT.
DR EMBL: D90854: BAA16050.1: ALT_INIT.
DR EMBL: K02672: ? NOT_ANNOTATED_CDS.
DR EMBL: U30459: AAA74094.1:
DR EMBL: Y00544: ? NOT_ANNOTATED_CDS.
DR Ecogene: E612850: yfal.
KW Hypothetical protein; Repeat: Signal.
FT SIGNAL 1 23
FT CHAIN 24 1250
FT DOMAIN 919 948
FT CONFICT 28 30
FT CONFICT 40 40
FT CONFICT 65 66
FT CONFICT 431 431
FT CONFICT 433 434
FT CONFICT 478 478
FT CONFICT 773 773
FT CONFICT 853 853
FT CONFICT 923 924
FT CONFICT 948 994
SO SEQUENCE 1250 AA; 131152 MW; 17F98C05E299FC95 CRC64;

Query Match 4.3%; Score 220.5; DB 1: Length 1250:
Best Local Similarity 22.7%; Pred. No. 5,9e-06;
Matches 143; Conservative 84; Mismatches 235; Indels 167; Gaps 26:

Oy 57 YLDNRLRIILLOKTPREGAAYT-----ITDYSFFPTQREGIV 95
Db 10 YLSLSPMISL-PSAGVAAYVDSGCGYVKASQASRSLSGITODMSIAGQ--WLV 66
Oy 96 FAKNLTPESGALGAYSPNSPTVEIRDTICPVIFENNTCCRPFTSSNPAAVKKIREGGA 155
Db 67 FSDMTNNAAGVAFLGQAGFSLPENEETGMTLFANNT-----VTGEYNNGA 114

```

```

Db 807 SADNTGTLFVNTPDI---TVTLNKGAGFEGVLKOVIIISGQNIYFENICNGVIGHGIA 863
Oy 248 SPICSLTNGRCNINIVNNRCFKNVETASSEASDC-----GAIKVTRLDVYGNRGRI-- 299
Db 864 NSISEMNSLGTSLFLPGSTPLDVLTIKSTYONGCVDFNAPIVVSGIDSMINNGQILIG 923
Oy 300 -----FSDN-ITKNGGAIYAPVTVLVNCPYFINNIANNKGAIYIDGTSNS- 348
Db 924 DKNIIALSLSDNSGSIYV-ANTLYSGIRTKKNNOGTVTLSCGMNNPGTIIYGLGLENGS 982
Oy 349 -KISADRRAIIFNE--NIVTVNTNAN-----GTSISANPRRRAITVASSGE- 393
Db 983 PKLKOVFTTDDYNNLGSITANNVTINDYVTLTGGIAGIDEDAK-----ITLGSVNGNA 1036
Oy 394 -----ILLAGSSQNLIFYDPIEVSNA-----VSSEFKADDTGSVY 432
Db 1037 NVRFVDSFSDPRSMIVATQAKGTVTYLGNAIYVNSIGSLDTPVASVRRF--GNDSCAGL 1094
Oy 433 FSGATVNSADPHQRLQTKTPAPLTLNGLFCIEDHAQLTVNRFQTGGVSVLSGNAVLS 492
Db 1095 QCNIVSQNIDFGTYNL-TLINSVILGGGTTAINGEIDLTTNNLIIFANGTSTWGDNTSIS 1153
Oy 493 CYKNAG-----NSASNASITLK-----HIGLMSLSIKGAEIPLLM 530
Db 1154 TTLNVSSGNIGOVVIAEDQAVNATTTGTTIKIODNANANSGTOAYTLIGGARF---- 1209
Oy 531 VEPTNNSNNYATDAATFSLSDVKLSLIDD-----YGNSP 565
Db 1210 -----NGTLGARNFVATGSIIFVKYELLRDSNODVYLTFTNDVNLVYTTAVGNSAIANAP 1264
Oy 566 YESTDLTHALSSQPMLSISEASDNOLRSDMDPSGLNPHYGMOGLTWGMAKTODPEPA 625
Db 1265 GVSONISRCLES---TNTAAVNNMLAKPDSVATFV-----GAITATDTSAAV 1309
Oy 626 SSATITDPPKARFHTLLTTLMLPAGVSPKHSPLNLTLMGNMLLTFSLKNSAEIL 685
Db 1310 TVVNLMDTKTOD-----LLSNRLGTLRLSNAETSDVAGS---ATGAVSSDEDE 1357
Oy 686 PS---DHPFMCITGGGLGMVYODPRENHPGFMHRSQVSGM--IAGOTHTFSLKFSQ 739
Db 1358 VSYGVAKKFFYN-----AEODKKGIAGAKKATTVGVGLDTLADNLMTGAIIGI 1409
Oy 740 TYTKL-NEHYAKNNVS-----KNVSCQEMLEFSLDEGLTKLVGLYSY 783
Db 1410 TKTIDKHODYKKDKTDINGLSFSLYSQQLVKNFPAQCNALFTTLNK-----V 1457
Oy 784 GDHNCHEFTTQGENLTSQ---GTFRSQTMGAVFF--DLPMPFGSTHILTAFLCALGI 838
Db 1458 KSKSQRYFFESNGKMSKQIAAGNYDMTFCGNLIFGYDYNAMP---NVLVTPMAGLSTL 1513
Oy 839 YSLSHFTEVG---AYPRSPSTPLINLVLPVIGVKSFEPMNAT 878
Db 1514 KSSNENYKETGTTIVANKRINSKFSRVDLIVGAKVAGSTVNT 1556

RESULT 5
OMP8_RICTY STANDARD: PRT: 1645 AA.
ID OMP8_RICTY STAN:
AC P96969;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE (SURFACE MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE MEMBRANE ANTIGEN) (CELL SURFACE ANTIGEN) (SCAS) (POMP B)
DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE).
DE SLP.
OS Rickettsia typhi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=785;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=WMILMINGTON;
RX MEDLINE=94040787; PubMed=8224886;
RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;
RT "Cloning and sequence analysis of the gene encoding the crystalline
RL Gene 133:129-133(1993).
RN [2]
RP PARTIAL SEQUENCE.
RC STRAIN=WMILMINGTON;
RX MEDLINE=92114896; PubMed=1370573;
RA Ching W.M., Carl M., Dasch G.A.;
RT "Mapping of monoclonal antibody binding sites on CNBr fragments of
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT prowazekii."
RL Mol. Immunol. 29:95-105(1992).
RN [3]
RP IDENTIFICATION OF CLEAVAGE SITE.
RX MEDLINE=92104668; PubMed=1729180;
RA Hackstadt T., Messer R., Cleplak W., Peacock M.G.;
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
RT membrane protein of rickettsiae: identification of an avirulent
RT mutant deficient in processing."
RL Infect. Immun. 60:159-165(1992).
CC -I- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -I- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A RICKETTSIAL
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC -----
CC EMBL: L04661; AAB48987.1;
CC DR Antigen; S-layer; Transmembrane; Cell wall.
CC K1 1353 120 KDA SURFACE-EXPOSED PROTEIN.
CC FT CHAIN 1 1353 32 KDA BETA PEPTIDE.
CC FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).
CC FT CONFLICT 657 657 H -> N (IN REF. 2).
CC FT CONFLICT 842 842 V -> I (IN REF. 2).
CC FT CONFLICT 1071 1071 G -> A (IN REF. 2).
CC FT CONFLICT 1306 1306 G -> S (IN REF. 2).
CC SO SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185EE CRC64;

Query Match 3.8%; Score 191.5; DB J; Length 1645;
Best local Similarity 19.9%; Pred. No. 0.00053;
Matches 206; Conservative 124; Mismatches 431; Indels 275; Gaps 44;
Oy 8 FLIGNSLSGL-AREVPSRFLFMPNSVPPDPKESLSKISITGTHNLTCYDNLRYILA 66
Db 629 FNVGSSKTLINAGDVAINELWMEND-----GSVHLTHNTYL----- 664
Oy 67 ILQKTT---PREGAAVITTDVLSFDFQKEGIVFAKNLTPESGGAIGVAPNSPTVEIRPT 123
Db 665 -IRKTTNANQGIITYAADPINDTALDQTNLSAESPLSNHFKATKANQSTI-LHIG 722
Oy 124 IGVIFENNTCCRPFTSSNNAVANKIREGAI-----HAQNLVYNH---NHQVGF 172
Db 723 KGVNLVANN-----ITTDANVGLSLHFRSGTISVGTGCGOGLKLNLLIDNGTIVKF 777
Oy 173 MKNFSVYRGAIISTANTFVSEN-----QSC-----FLFMDNICIQNTAGKGAIV 219
Db 778 LGDITFNGGFKRIGKSLIOJISSNYITDHTIESADNTGTLFVNTPDI---TVTLNKGAGF 834
Oy 220 AGTSNSFESNNCDLFF--INNACCAGAIKPSPICSLTGNNGNIVFYNNRCFKNVETASSE 277
Db 835 GVLKQVWVSGPNIATFRIEIGNVAAHAIAYDS-ISFENASIGASLFLSLGTPLDVLTIKST 893

```

```

OY 256 NRGNIY-----FYNNRCFKNVETAS-----SEASDGAIAVTRLDVYGNRGRI--- 299
DB 532 DAGNTMPLTIKSTVGNKTAKEGDPVSVVVLGVDSVIADG-----QVIDQNNIYGL 582
OY 300 -FFSDNITKNNGGAIYAPVVTLVNDGPTFYNNIANKNGGAIYIDGTSNKSISADRHAI 358
DB 583 GLGSDNGIIVNATTLVAGISTLNNNOGTVTLGSGVPNTPGTYVGLTG---IGASK---- 635
OY 359 FNEIVTNTVANGSTSNAPPRNAITVASS-----SGEILGA-GSSONLIFYPDI 410
DB 636 FKQVFTTIDYNNLGNIIATNATINDGVYTTGGIAGIGDGIITLGSVNGNNGVRFADI 695
OY 411 EVSNAGVSVSFNKADOT-----GSVFSGATVNSA----- 441
DB 696 -LSNSTMIGTKKANNGVTYLGNAFVGNIGSDPFVAVSVRFSGSDSGAGLOGNIYSQVI 754
OY 442 DEHQNRLOTKTAPRLTSLNGFLIEDHQAOLVNRFTQGTGVVSLNG---AVLSCKYKNG 497
DB 755 DFGTYNLGI-VNSNIIIGGTTAINGKIDLVNTLTFASGTSTWGNNTSIETTLTANGN 813
OY 498 AG-----NSASNASITLK-----HIGLNLSSILKSGAEI-----PLLWEP 533
DB 814 ICHIIYLECAQVNTTTTGTITTIKVODNANANFSGTOTYTLIOGAFNGLTSGPNFVVG 873
OY 534 TNNSNVYTAATAAT-----FSLDYKLSLIDYGNSPY-----ESTDLTHAISOPM 580
DB 874 SNRFVYVSLIRANQDYVITRTNNMENVTNDIANSFGGAPGVQDQNTFVNATITAY 933
OY 581 LISSEASDNLQSDDMDESGLVNPHYGWGLTWGAKIQDEPPASATITDPO----- 634
DB 934 NNLLAKKN-----ANSANFVGAIVTD-----TSAITVYQDLKARD 970
OY 635 -FANRHFLLLTWLDAGVVPSPKHSPLIANTMGNMLLATESKNSAELTPSDHPFG 693
DB 971 IQAQQLNRLGALRYLG---TPETAEMAGPEACAIANAANGDAIDQNA-----YG 1018
OY 694 ITGGGLGMVYODPRENHQCFHMRSSGYSAGM--IAGQTHFSFLKFSQTYTKL-NERYAK 750
DB 1019 IMAKPPTYDAHOGKKGGLAGYKAKTGVYIGDLTDLANDMLCAIGITKTDIKHODYK 1078
OY 751 NNVS-----KNYSCGEMLFSLQEGFLLTKLVGLXYGDHNCNHHIYTOG 795
DB 1079 GDKTDVNGFSFSLYGAQOLKKNFQAQSAIFSLNQ-----VKKRSQRYFFDAN 1126
OY 796 ENLTSG---GTFESQTMG--AVFEDLPKMPFQSTHILAPFELGALGIYSSLSHFTEVGA 850
DB 1127 GNMKSQIAGHYDMTFGGLNLYGIDYN---AMOGVLYTPMAGLSYLKSDENYKETGT 1182
OY 851 -----YPRSFSTKTPLLINLVPIGVGSFMATQRPQAWTVELAYQPVLYR----- 896
DB 1183 TVANKQVNSKFSRDTDL---VGAKVAGSTMNT-----DLAVYREVAHFVHKVTG 1231
OY 897 -----OEGCITQLLASKGIMFGSGSPSRHAKSKIKSQOQPLSMLT 939
DB 1232 RLKSTQSVLDGOVTPCINOPDRITKTSYNLGL---SASIRSDAKMEGXIGYDAQISSKYT 1288
OY 940 LH 941
DB 1289 AH 1290

```

RESULT 7

```

OMPB_RICRI STANDARD: PRT: 1654 AA.
AC 053047;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCAS) (ROMP B)
DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE).
GN OMPB.

```

```

O5 Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=783;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-R;
RX MEDLINE=92167802; Pubmed-1724278;
RA Gilmore R.D. Jr., Cleplak W. Jr., Policastro P.F., Hackstadt T.;
RT "The 120 kilodalton outer membrane protein (Romp B) of Rickettsia
RT rickettsii is encoded by an unusually long open reading frame:
RT evidence for protein processing from a large precursor."
RN Mol. Microbiol. 5:2361-2370(1991).
RL (2)
RP SEQUENCE OF 279-1654 FROM N.A.
RC STRAIN-R;
RX MEDLINE=90136087; Pubmed-2515418;
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT "Cloning, expression and sequence analysis of the gene encoding the
RT 120 kD surface-exposed protein of Rickettsia rickettsii."
RL Mol. Microbiol. 3:1579-1586(1989).
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X16353; CAA34403.1;
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1333 120 KDA SURFACE-EXPOSED PROTEIN.
FT DOMAIN 1334 1654 32 KDA BETA PEPTIDE.
FT DOMAIN 1181 1188 POLY-THR.
SO SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;

```

Query Match

Best Local Similarity 19.4%; Score 190; DB 1; Length 1654;

Matches 198; Conservative 130; Mismatches 424; Indels 270; Gaps 42;

OY 92 EGI-YFAKNLTPESGCAIGASNSPTVEIRDTIC-----PIVFENNTCCR-----P 137

DB 721 EGVNLVATNTTTTDANVGSVFNAGGTINIVSGTVGGQCKKFTVALENTYKFLGNAT 780

OY 138 FTSSNPAAVANKIREGGAIAHONLIYINHNHDVGFPMKNEFVYVGAISTANFVSENO 197

DB 781 FNGNTTIAANSTIQIG-----NTADCYA-----SADGTGIVEFNTGPITVTLN 826

OY 198 CPLEMDNICIQITNACKGCAIV--ACTSNSFESNCDLFFINNACGACGAIISPISLTG 255

DB 827 KQAPVALKQITVSGPGNVVINEIDGNAGNHGAVDTITAFENS-SLGAVVFLPRGIPFN 885

OY 256 NRGNIY-----FYNNRCFKNVETAS-----SEASDGAIAVTRLDVYGNRGRI--- 299

DB 886 DAGNTMPLTIKSTVGNKTAKEGDPVSVVVLGVDSVIADG-----QVIDQNNIYGL 936

OY 300 -FFSDNITKNNGGAIYAPVVTLVNDGPTFYNNIANKNGGAIYIDGTSNKSISADRHAI 358

DB 937 GLGSDNGIIVNATTLVAGISTLNNNOGTVTLGSGVPNTPGTYVGLTG---IGASK---- 989

OY 359 FNEIVTNTVANGSTSNAPPRNAITVASS-----SGEILGA-GSSONLIFYPDI 410

DB 990 FKQVFTTIDYNNLGNIIATNATINDGVYTTGGIAGIGDGIITLGSVNGNNGVRFADI 1049

OY 411 EVSNAGVSVSFNKADOT-----GSVFSGATVNSA----- 441

01-NOV-1997 (Rel. 35; Last annotation update)
DE VACUOLATING CYTOTOXIN PRECURSOR.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95355366; PubMed=7629077;
RA Atherton J.C., Cao P., Peek R.M. Jr., Tummuru M.K., Blaser M.J.,
RA Cover T.L.;
RT "Mosaicism in vacuolating cytotoxin alleles of Helicobacter pylori.
RT Association of specific vaca types with cytotoxin production and
RT peptic ulceration."
RL J. Biol. Chem. 270:17771-17777(1995).
CC - FUNCTION: INDUCES VACUOLATION OF EUKARYOTIC CELLS. CAUSES
CC ULCERATION AND GASTRIC LESIONS.
CC - SUBCELLULAR LOCATION: SECRETED.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U29401; AAA66834.1; -
DR CYTOTOXIN; Toxin; Signal.
FT SIGNAL 1 30
FT CHAIN 31 310
FT PROPEP ? 1310
FT SEQUENCE 1310 AA: 141988 MW: 18C21FE3D435F981 CRC64;
Query Match 3.4%; Score 173; DB 1; Length 1310;
Best local similarity 21.2%; Pred. No. 0.0054;
Matches 225; Conservative 137; Mismatches 374; Indels 360; Gaps 60;
QY 7 FLLIGSL-SGLAREVPSRIELPNSVPDPTKSLSKSL-TGDTHTLNTN----- 55
DB 209 FVEINRVGSGAGRKASSTVLTKSEKITSRENA--EISLYGATLNLVSSNOSVDLY 266
QY 56 --CYLNLRLYLALLOKTPREGAAVITIDVLSFPDQK--EGIFAKNL-----TPESGA 107
DB 267 GKVMGRLGVAGYVLAPS-----YSTIDTSKVOGEMNFRHLAVGDONAAOAGI 314
QY 108 IGYASNPSPVEIREDTIGPYI-----FENNTCRPFSTSSNPNAVYKIR----- 151
DB 315 IAKKKNITIGLDMQSGSLITIPREGYSEKTKDNQNPKNDAQTEIQPOVIDPFE 374
QY 152 EGGAIHQNLY-INNHND---VVGPMKNFS-----YVGGALSTANTVSENOGSCFL 200
DB 375 AGGKDTVNFHNTKADGTLRAGGFKASLSTNAHLHIEGGVNLN-----QASGRTL 429
QY 201 FMONICIGTGTACK-----GGAIYAGTSNSPE-----SNNOCLPFTNNCAGCAIF 247
DB 430 LVEMLTGNTVEGTLRVNNOVGGALAGSSANEEFKAGEDTNNATATF--NNDIHLKAVN 488
QY 248 SPICSLTGN-RGNI-----VFYNNRCFKNVTASSE----- 277
DB 489 LRDADHTANFNKNIYLSKSTNLKRNKHTAFKKNIDATKSONGLTSTLDPSGVTDKVNIN 548
QY 278 ---ASDGAIAK-----VTRLDVTGNRGRIFESDNI-TKNYGAIA-----YAP--- 316
DB 549 KLTAAATNVIKKNEDIKELVYTVTFVQSGO--YTFEENIGDKSRIGVSLQTSYSPAYS 606
QY 317 -----VTVLVNGP-TYF-----IN-----NINNNKGAIYIDGTS 346
DB 607 GGVTFKGKKLVLDIETIHAPNIFDARNVTDVETNKRILFEGAPGNIA-GKITGLMFNNLTL 665

QY 347 NSKISDRHAIITFENIYVNTNANGTSTSANPPRNATVYSSSGEIT-LLGAGSSQNL 405
DB 666 NSNWSMD-YGKDLDTLIDGHFTNNQCT-----MNLFEVODRVAITLWAGQASMI 713
QY 406 -----FYDPIEVSNAGVSVFKNKADQTS-----VFESGATVNSADFHQRML 448
DB 714 FNNLUDSTGTGFKPLIKINNAQNLTKNKEHVLVKANRIDNVLGVGASDYDINSASNTNL 773
QY 449 QTKTPAPLTLNSG-----FLCI--EDHAQOLTVNRFOTQGVSLGAVLS----- 492
DB 774 QEPFKRLALYNNNNMDTCVVRKON-----LNDIKACG--MAIGNQSMVNNPENYKYLEG 827
QY 493 -CYKN-GAGNSASMSITTKHIGLNLSSILKSGAEIPLLVNVEPTNNSNNTADTATFSL 550
DB 828 KWKKNKGIKKTNANTTIAV-NLG-NNSTPTNSTDTITNL--PTNTNN-----ARFA- 875
QY 551 SDVKLSLIDYGNPY--ESTDLTHALSSQPMLSIASEADNQLRSDMDPFGSLNVPHYGM 608
DB 876 ---SYALIK--NAPFAHSATPVLVAIINHDEGTIESVELANRSSDIDITLVANSAGOR 929
QY 609 OGLMTW-----GNAKT-----QDEPPASATITDPOKANRFRHLLTLWLAG-- 651
DB 930 DLQOTLIDSHDAGVARTMIDATSAEITQOLNAATTLNNIASLEBKTSGLQTLISLNA 989
QY 652 -----YVSPKHSPLIANTLM 668
DB 990 MINSRLVNLRRKHTNHIDFAKRLQALDORFASLESAEVLVYOPAPYERK--TNW 1046
QY 669 GNNLMTESLKNSAEILPSPDPFWGITGGLGMVYODPRENHPGHRMSSGACMIAG 728
DB 1047 ANAIGCT-SLNG-----SNASLYG-TSAGVDAYLNGEVAIYGGF--GSYGRSS----- 1092
QY 729 OFHTFSLKESQTYTKLNERYAKNN--VSSKNYSQGGEMLFSLQ-----EGFL 774
DB 1093 -----FSNQANSLNSGANNNTNFGVYSRIIFANQHEPDEAGALGSDQSLNFKSALL 1144
QY 775 TKLVGLSYGDHNCHEFTYOGENLTSOGTFRSOTMKGANVF--DLPMKP-----FGS 824
DB 1145 QDNLQSYHY-----LAYSATTPRASGYDFAFRNALVLKPSVGVSYNHLGS 1190
QY 825 THILTAFLGALGIYSSLSHFEVGA 850
DB 1191 TNKNSNOVALSNGSSOHLFNANA 1216
RESULT 12
Y589_CAEEL STANDARD; PRT: 2329 AA.
ID Y589_CAEEL
AC 009624;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHEICAL 254.3 KDA PROTEIN ZK945.9 IN CHROMOSOME II.
GN ZK945.9
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Wilkison-Sproat J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

Query Match 3.2% Score 161; DB 1; Length 1286;
 Best Local Similarity 20.6% Pred. No. 0.029;
 Matches 219; Conservative 115; Mismatches 359; Indels 370; Gaps 61;

```

Oy 69 OKTRPEG-AAATITDYLSFDFDOKECIYFAKMLTPESGAI-----GYASPSPIVEIRD 122
Db 349 OHIRNGIAGSIATVINGSVNISGG--YAESTIINSGLTSLVSDYA-----RG 397
Oy 123 TIGPVIFFENNTCCRFPTSSNPAAVN-KIREGCAIHAOLYINHHNDVYGFKNFSYVR- 180
Db 368 TI-----LMSRE--NMSGVSYNAMINTG-----NOIYSOCEATAIVNTSGOR 445
Oy 181 ---GGAISTANTFVVSSENOSEFLFMDNICIQTNTAGK--GGAIVAGTNS-----FES 228
Db 446 INSGTAPVQNSVVTFRVVS-----SAKPFDAEVYSGKQTYLWLMGIWTS 492
Oy 229 NNCDLFFINNACCAGAIFFPICSLTGRNGNIV--FYNNRCKNVEPTASSEASDGAIKV 286
Db 493 NFLTAVMSNFPCTAGANVNLGRLNAFAGNVYGTILNDEGQVYVSGATATSTVG--- 548
Oy 287 TRLDVTGMRGRIF-----FSDNITKNYGG--AIYA---PVYTLVNDGPTVEI----- 329
Db 549 -----NNEGEYVLSGCTIDGTVLNSGGLAVSSGKASATVINEGAQVYVYDGGQVT 601
Oy 330 -NNIANNNKGAIIYID-GTSNSKISADRAHAIIF-----NENIVTNTN 369
Db 602 GTNININ--GGTIRVDSCASALNIALISSGGNLTSTGATLPELTMAALSVSNHMSNIYL 659
Oy 370 ANG-----TSTANPPRR-----NAITYASSGCEILLGAG--SSONLIF- 406
Db 660 ENGGCLRYTSGTADDTITVNSAGRLRIDDGTINGTITINAGC-IVAGNININDGFIIN 718
Oy 407 ---D-PIEVSNAGVSFVN-----KEAD---QTSVVSFGATVSADPHQN 447
Db 719 LAENYDFEELSESGVLYVNDNTGIMTYACTLYOAGCVNKNKGIIIFDSAVVN-ADM---- 773
Oy 448 LOTKPAPLTLSNGELCIDHAQLTVNRTOTGVVSLNGAVLSCYKNGAGNSASNASI 507
Db 774 -----AVNQNAVINISD--QATINGSVNNNGSVIVNNSII-----NONTITDADL 816
Oy 508 ---TLKHIGLNLSSILKSCAEIPLWVEPTNNS-----NNTADTATFESLSDV---K 554
Db 817 SEGTAKLISATVNSGLVNNKNIIL--NPTKESAGNTLVSVYTGTPGSVLSLGGLEGD 873
Oy 555 LSLIDY---GNSPESTDLTHALSQPMLSISEASDNLR-----SDDMDFSL 601
Db 874 NSLTDLRVKNGTSGOS-DIVYV-----NEDSGGOTROGINIISVEGNSDAEFSLK 924
Oy 602 N---VPHYGM-----OGLMTWGNAKTOD-----PEPASSATITDPOKANRFRH 641
Db 925 NNVAAGAYITLQKNGESTDNKGMWLTSLPTSDROYRPEMGSAVT----- 972
Oy 642 TLLLTWLPAGVPSRKHNSPLIANTLMGNMLLATESLKSASBLT-PSDHPFW-GITGG-- 697
Db 973 -----NMALANSLEFLMDLNRKOFRAKMSDNTOPESASVMMKILITIGIS 1014
Oy 698 -----GLGMVYODPRENHPGFHMRSSG--YSAGMIMGOTHTFLKFS 738
Db 1015 SKRLNDGNKTTTNOPIINGLDGIYK-----FHAEDLGDTLGLMGVYANAKCKTIN 1066
Oy 739 QTYTKLNERYAKNNVSSKNYSCOGEMLFSLQEGFLTLTKVLGLYSYSDHNCHEFTYGENL 798
Db 1067 YTSNKA-----ANRTLD--GYS-----VCVYG-----TWYONGENA 1095
Oy 799 TSGCTFRSGTMCAGAFEDLPKMPFGSTHILTLAPLGCALCIYSSLSHFTVGAIPNSFSFK 858
Db 1096 T--GLF-AETWQYNNMFNASVSKDG---LEEEKYNLNLITASAG-----GGYNLWVHW 1143
Oy 859 TPLINVLVIGVGSFMANNTORPOAWTVELAYOPLVYROBPGLATOLLSKGIWFGSGSP 918
Db 1144 TS-----PEGITGEFW---LOPHLOAVMMGVTPTHQEDNGTVVODACKNNIQTACIR 1194
Oy 919 SS-----RHMSYKISQOTOPLS 936

```

```

Db 1195 ASWKVSTLKDXTGRFRFPYIEANNIHTHEGVKMSDQSLS 1238
RESULT 14
PK2-DICDI
ID PK2-DICDI STANDARD: PRT: 1858 AA.
AC P54674;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PHOSPHATIDYLINOSITOL 3-KINASE 2 (EC 2.7.1.137) (PI3-KINASE)
DE (PDIINS-3-KINASE) (PI3K).
GN PI3K OR PIK2.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=96009592; PubMed=7565716;
RA Zhou K., Takegawa K., Emr S.D., Firtel R.A.;
RT "A phosphatidylinositol (PI) kinase gene family in Dictyostelium
RT discoideum: biological roles of putative mammalian p110 and yeast
RT Vps34p PI 3-kinase homologs during growth and development.";
RL Mol. Cell. Biol. 15:5645-5656(1995).
CC -I- CATALYTIC ACTIVITY: ATP + 1-PHOSPHATIDYL-ID-MYO-INOSITOL = ADP +
CC 1-PHOSPHATIDYL-ID-MYO-INOSITOL 3-PHOSPHATE
CC -I- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U23477; AAA85722.1; -.
DR DictyDB: DD01100; PI3K.
DR InterPro: IPR000341; -.
DR InterPro: IPR000403; -.
DR InterPro: IPR001263; -.
DR InterPro: IPR002420; -.
DR pfam: PF00792; PI3K_C2; 1.
DR pfam: PF00794; PI3K_Rbd; 1.
DR pfam: PF00613; PI3K; 1.
DR pfam: PF00454; PI3-PI4_Kinase; 1.
DR PROSITE: PS00915; PI3_4_KINASE_1; 1.
DR PROSITE: PS00916; PI3_4_KINASE_2; 1.
DR PROSITE: PS50290; PI3_4_KINASE_3; 1.
KM Transferase; Kinase; Multigene family.
FT DOMAIN 34 40 POLY-GLY.
FT DOMAIN 166 172 POLY-SER.
FT DOMAIN 185 226 POLY-ASN.
FT DOMAIN 227 235 POLY-THR.
FT DOMAIN 246 253 POLY-SER.
FT DOMAIN 259 268 POLY-ASN.
FT DOMAIN 294 303 POLY-THR.
FT DOMAIN 361 364 POLY-SER.
FT DOMAIN 369 384 POLY-SER.
FT DOMAIN 425 429 POLY-GLN.
FT DOMAIN 439 444 POLY-THR.
FT DOMAIN 445 454 POLY-SER.
FT DOMAIN 562 570 POLY-GLY.
FT DOMAIN 715 727 POLY-THR.
FT DOMAIN 982 990 POLY-GLN.
FT DOMAIN 1015 1049 POLY-ASN.
FT DOMAIN 1598 1858 PI3K/PI4K.
SO SEQUENCE 1858 AA; 203945 MW; A6C033304CDEAD03 CRC64;

```

Query Match

3.1% Score 159.5; DB 1; Length 1858;

```

Db 452 IGVA-----GVCN-----LNTSDCKEYSON-----ITFLGKAGSIGTL 486
OY 439 NSADFHQNLQTKTPAPLTLNGFLICIEDHQAOLTVNRFOTGVVSLGCAVLSCKYNG- 497
Db 487 NLMD-----ATSSF-----DYGVINGNF--GSGIVNNGCATLNTGTGFG 525
OY 498 ACNSASNAITLKHIGLNLISLTKSGAEPLIWEPTNNSN-----NYRAD 543
Db 526 IGNASNGKI-----VNIST-----DSLMLNKTSTNQLLOVGLTGEINLT-- 569
OY 544 TAATSLSDVKSLLD-----DYNSPYESTDLTHALSOPPLSLISEADNOLRSD 595
Db 570 TGGIVKARDTQIALNDKSKGVDRVQDQNSLLETFMNVGTSGTGLTLTNNGTFLNVEGGE 629
OY 596 MDESLANPHYG-----MOGLWTWGMAKTQDEPPASS 627
Db 630 V-YLGVFEPVAVCTLNIGAHGEAADAGTITNATVEFGLGCVFVFNHTNNSDAGQYVD 688
OY 628 ATITDPOKANRFHRTLTLTWLPAGYVVPSPKHSPLIANTLMGNMLLATESLKSAPLPS 687
Db 689 MLITGDDKDKYIHDAGHTVFNAG-----NTYSG-----KTLVNDGLTLTA 729
OY 688 DHPFNGITGGGIGMMVYODP-----RENHGFH-----MRSSGSAGMIAGQ 729
Db 730 SHTADGVYGMGSEVYTIAPGTLDILASTNSAGDYTLTNALKGDLMRVQLSSDCKMFGF 789
OY 730 TTTFSLKFS-----QYTKLNERAKNNVSSKNVSCOGEMLFSLQEGFLTKVLGLYS 782
Db 790 TIAATGEFAGVQLKDSTTT-----LERNTAALHTA--MLQSSSEVTTGVK--VGEOS 839
OY 783 YGDHNCHEFTYTGENTSGCTFRSQTMGAVFF--DLPMKPFSGTHIILAPPLGALCIYS 840
Db 840 IGLAMN-----GCTIIFDIDIPATLAEGIVSDTIVVGADYT 879
OY 841 SLSHFEVCAVRSFSTKPLINLVPIGVKSGFPMATROPAVYELVOP-----VLX 895
Db 880 MKGRNYVN-----GTGDVLIDY-----PKPMDPMANNPLTTLNLE 917
OY 896 ROEPGIATOLLASKGIWFGSGSPSRHANSYKISQOTQPLSLMLTLPFYHG 946
Db 918 HDDSHVGVLVKAKQYVIGSGSLTLRLDLOGDEVDK-----TLHIAONG 962

```

```

RESULT 16
YFJA_ECOLI STANDARD: PRT: 1569 AA.
ID YFJA_ECOLI
AC P52143: P76610: P77017: P77019:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHEICAL 98.4 KDA PROTEIN IN ALPA-CABD INTERGENIC REGION (F949).
GN YFJA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxId:562:
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RP SEQUENCE FROM N.A.
RC STRAIN-K12:
RA Alba H., Baba T., Fujita K., Hayashi K., Honojo A., Horiuchi T.,
RA Ikemoto K., Inada T., Isono K., Isono S., Itoh T., Kanai K.,
RA Kasai H., Kashimoto K., Kim S., Kimura S., Kitagawa M.,

```

```

RA Kitakawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H.,
RA Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Oshima T.,
RA Saito N., Sempel G., Seki Y., Tagami H., Takemoto K., Wada C.,
RA Yamamoto Y., Yano M.;
RL Submitted (Jan-1997) to the EMBL/GenBank/DBJ databases.
CC - SUBCELLULAR LOCATION: OUTER MEMBRANE (POTENTIAL).
CC - SIMILARITY: STRONG. TO BORDETTELLA PERTACTIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: U36840: AAA79815.1: ALT_SEQ.
DR EMBL: AE000350: AAC75695.1: -.
DR EMBL: D90889: BAA16514.1: ALT_INIT.
DR EMBL: D90890: BAA16518.1: ALT_INIT.
DR Ecogene: EG13213: yfja.
KW Hypothetical protein; Outer membrane.
SQ SEQUENCE 1569 AA; 162774 MW; 0D006A9A32154596 CRC64;

Query Match 3.0%; Score 155; DB 1; Length 1569;
Best Local Similarity 19.7%; Pred. No. 0.088;
Matches 194; Conservative 125; Mismatches 320; Indels 344; Gaps 51;

OY 97 AKNLPESCAIGAYASPNPVEIRDTIGPVIEENNTCCRPSTSPNPA-----AVN 148
Db 579 AKHVEQSGCALIASTTSGTILIEGTNSYDARYIRNSEAKNVYLENAGSLVYVTSRAVD 638
OY 149 KI-----REGAI-----HAQNLXNNHND----- 168
Db 639 TIINAMCKMDYGVKQDVGYLNSAGTQTIYASATSDKANIKGKQVYGLATEANIESGEO 698
OY 169 -VGFPMKNSYVRGCAIS-----TANFVSENSCFLFMDNICIQTTAGKCAIYNG 221
Db 699 IYDGGSTEXTHINGTQTOYQNGKAIKNTDIVSGLQ-----IMAGTAEGLIING 749
OY 222 T-----SNSFESNCDLFFINNACCAG-----GAIFSP--TCSLTGRNNGVIFV- 263
Db 750 SOYVNEGGIAENSVLNDGGLDYREKGSATGIGQSSQGLAVATRTATRTGTRADSVAS 809
OY 264 -----NNRCFKN--VETASSEASD-----GGAIKVTRLDVYGN-----RGR 299
Db 810 IEGGANNILLANGLVLYESDTSQKTOVNMGGREIVTKATATGTLTGGEQIVEG-- 867
OY 300 FESDNTIKYGC-----AIYAPVYLVDNGPYFI-----NIANNKCAIYIDGTSNS 348
Db 868 -VANETTINGDIQIVSANGAEAIKTINGGTLVDNKGKATDIYONSSAALDTSTANI 926
OY 349 KISADRHAIIF--NENIVNVTVANGSTSANPRRNATIVASSG--EILIGASSONLI 405
Db 927 EISGTHQYGTFSISGNLATMMLLENG-----GNLVLAEATKARDSTVKGGMQNI 977
OY 406 FYDPIEVNAGVSVSFNKDAQ-----TGSVFESG-----ATVNSADFHQHN 447
Db 978 GQDSATKVNKSGGYTLGRSDEQALARAEDLQVAGCTAIYVAGTILADSVSGA--TGS 1034
OY 448 LQTKTP---APLTLNGLICIEDHQAOLTVNRFOTGVVSLGCAVLSCKYNG- 497
Db 1035 LSLMTPRDNTVPKL-EGAVRITDSATLT-----LQNGVDITLADITLASSRS 1081
OY 498 -----ACNSASNAITLKHIGLNLISLTKSGAEPLIWEPTNNSNMYVDTATFSLSDV 553
Db 1082 VMLNSNNSCAGTS-----NCEYRVNSLNDGQDYL-----SAQTAAPATTNGI 1125
OY 554 KLSL-----IDDYNSPYESTDLTHALSOPPLS-----ISEADNOLR-- 592
Db 1126 YNLTJTNELSGSNF-YLHTNVAGSRGDDQLVYNNNNTGNKIFVQDPTGVSPQSDAMTLV 1184

```



```

RL EMBL: AE000291: AAC75061.1: ALT_INIT.
CC -1- FUNCTION: CONTROLS COLONY FORM VARIATION AND AUTOAGGREGATION. MAY
CC -1- FUNCTION: AS AN ADHESIN.
CC -1- SUBUNIT: HETEROOLIGOMER OF THE PERIPHERAL MEMBRANE PROTEIN (ALPHA
CC CHAIN) ANCHORED TO THE INTEGRAL OUTER MEMBRANE PROTEIN (BETA
CC CHAIN).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE ASSOCIATED.
CC -1- SIMILARITY: TO ADHESIN AIDI-1 AND TO BORDETELLA PERTACTIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000291: AAC75061.1: ALT_INIT.
CC EMBL: D90838: BAA15825.1: ALT_INIT.
CC EMBL: D90839: BAA15832.1: ALT_INIT.
CC EMBL: U24429: AAB47869.1: -.
CC Ecogene: EC12686: flv.
CC Outer membrane: signal.
CC SIGNAL 1 52
CC CHAIN 53 551
CC CHAIN 552 1039
CC VARIANT 2 2
CC VARIANT 41 42
CC VARIANT 46 46
CC VARIANT 157 157
CC VARIANT 188 188
CC VARIANT 303 305
CC VARIANT 320 320
CC VARIANT 372 372
CC VARIANT 433 433
CC VARIANT 497 497
CC VARIANT 585 585
CC VARIANT 709 709
CC VARIANT 721 721
CC VARIANT 751 753
CC VARIANT 803 803
CC VARIANT 815 815
CC VARIANT 824 824
CC VARIANT 829 835
CC VARIANT 845 847
CC VARIANT 855 855
CC VARIANT 888 888
CC VARIANT 1025 1025
CC CONFLICT 61 63
CC SEQUENCE 1039 AA: 106841 MW: 5170647C8DBE80 CRC64:

```

Query Match 2.9%; Score 147; DB 1: Length 1039;
 Best Local Similarity 20.6%; Pred. No. 0.16;
 Matches 200; Conservative 126; Mismatches 306; Indels 340; Gaps 53;

```

QY 146 AVKIKREGCA--IH-----AQNLYNNHNDVYVGMKNSYRGALISF--ANFVV----- 192
DB 146 AVNTTLNGGGMHMEGAIAIYVIND-----KGMQVVKPPTVADIVYVNGAEGGPD 197
QY 193 SENOSCFLEFMDNICIOTNTAGKGG-----AIYAGTSNFSNNCLF----- 234
DB 198 AEMGDGQGFVGRGAVR-TINKNGRQIVRAEGTANTTVVAGGDDQYHGHADLTTLNGGY 256
QY 235 -FINNACGAGAIIFS---PICSLTGNRGNIVFYNNCFKNVETASSEAS---DGAIR 285
DB 257 QYVHNGGTASDVTYVNSGQWIVKNGVAGNTT-VNCKGRGLVDAGGTATVTLKGGALV 315
QY 286 VTRRLDVTG-NRGRIF-----SDNITKNGCAL-----YAPVTVLVNDQPTFIYN--- 331
DB 316 TSTAATVGTINRLGAFVSEVKGADNVVLENGGRDLVLTGHTATNTATRVDDGGLTIDVRNGT 375

```

```

QY 332 ---IANNKGAIYIDGTSNSKISADRHAIENENITVNTANGSTANPPRRNATIVA 388
DB 376 ATTSMNGCVLLAD--SGAAVSG-----TRSDGKAFPSIGCGQADA----- 414
QY 389 SSSGELLGAGSSQNLIFDPIEVSNAGVSVFNKEADQTSVFSGATVNSADPHORNL 448
DB 415 ---LMEKGSSTL-----NAGDTA-----TDTWNGGLTTAKO- 445
QY 449 QTKPAPLTLSNGFLCIEDHAQLTVNRFTQ-----GGVSLGNCALVSCYKNAG 499
DB 446 -GLACTTLNNG--AIIULSKTYVNDLTIREDDALLQGGSLT-GNSV---EKSGG 498
QY 500 N-SASNASITLKHIGLINS-----ILKSGAEIPLW-----VEPTNSNN 539
DB 499 TLTVSNITLTKAVNLNECTLLNDSTVTTDVIAGRTALKLTGSLVNGAIDPT---N 554
QY 540 YTAQTAFTSLSD--VKLSLIDY---GNSPYEST-----DLTHALS 576
DB 555 VTLASGATWNIIDNATVQSVDDLSHAGIHTSTRTGKEVPATLKVNKNGONGISLR 614
QY 577 SPMPLISEASDNLRSDDMDFSG--LNVPHYG--WQGLMTGMAKTQDPPASSATIT 631
DB 615 VRPDMAONNA--DRLVYIDGRATGKTLNLVYVNGNSASGLATSG--KGIOVFAINGATTE 671
QY 632 DPO-----KANRPHRTLLTWLPAGVVPSP--KHRSPLIANTL-----WGNMLATE 676
DB 672 EGAFVQGNRLQAGAFVYSLNRDSESVYRSENAVRAEVLVYASMLTQAMDYDRIVAGSR 731
QY 677 SLK-----NSAEILPSHPFWGITGCGLCMMVYDPPENHPGFMHSSGVSAGAIQAO 729
DB 732 SHQTGVNGENNSVRLS-----IQQGHLG-----HDNNGGARGATPSSG----- 771
QY 730 THFSLKFSQTYTKLERYAKNNVSSKNYSKOGEMLSQEGFLTLKVLGYSGDHNC 789
DB 772 -----SYGFVRL-----EGLMRTVEAGMSVY--AGVYGAAGHS-- 803
QY 790 HFYTGENTLSQGTFRSQTMGAVFFDLPMKPPGSTHILTAFLPLAGLYSSHFTEVG 849
DB 804 SYVDKDDDSRACTVDDA-----GCLGGLYMLNVH----- 833
QY 850 AYPRSFSTKPLINLVPIG-----YKGFMMNATORQAM-----TYELAVO 891
DB 834 -----TSSGLMADIVAAGTRHSMKASSDNDNFRARGWGLGSLTGLPFSTIDNMLE 886
QY 892 PVLVROEPGIAT-----QLASKGIMPGSGSPSSRRHAM--S 925
DB 887 POLQTYWGLSLDDGKDNAGYKFGHGSQHYRAGRFLGSHNDMTGBGT--SSRAFLRDS 945
QY 926 YKISQQTPLSW 937
DB 946 AKHSVSELPVNW 957

```

RESULT 21
 SLAP-CAUCR
 ID SLAP-CAUCR STANDARD; PRT: 1025 AA.
 AC P35828: Q46015.
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE S-LAYER PROTEIN (PARACRYSTALLINE SURFACE LAYER PROTEIN).
 GN R5AA.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 OC NCBI_TaxID=76;
 RN (1)
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21; 646-660 AND 1021-1025.
 RC STRAIN-ATCC 19089 / CB15;
 RX MEDLINE=93007489; PubMed=1393820;
 RA Gluchrist A., Fisher J.A., Smit J.K.;
 RT "Nucleotide sequence analysis of the gene encoding the Caulobacter
 crescentus paracrystalline surface layer protein.";

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 19:16:45 ; Search time 71.24 Seconds
(without alignments)
1587.670 Million cell updates/sec

Title: US-09-677-752-2
Perfect score: 5086
Sequence: 1 MKKAFEEFLIGNSLGARL.....GFYSSPFNCYLNGLRIALRF 965

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

SPREMBL-15:*

1: sp-archaea:*

2: sp-bacteria:*

3: sp-fungi:*

4: sp-human:*

5: sp-invertebrate:*

6: sp-mammal:*

7: sp-mhc:*

8: sp-organella:*

9: sp-phage:*

10: sp-plant:*

11: sp-rodent:*

12: sp-unclassified:*

13: sp-vertebrate:*

14: sp-virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3648.5	71.7	976	2	09PL47
2	1120.5	22.0	938	2	092883
3	1060.5	20.9	952	2	09K2A5
4	1052	20.7	934	2	092882
5	1046	20.6	934	2	09JSE7
6	951.5	18.7	946	2	09RB60
7	948.5	18.6	946	2	092880
8	793	15.6	1025	2	09PL46
9	655.5	12.9	1013	2	084879
10	645	12.7	936	2	092898
11	644	12.7	936	2	09JSA2
12	636	12.5	926	2	09JSA2
13	620.5	12.2	987	2	P71135
14	620	12.2	930	2	09RB65
15	617	12.1	930	2	092393
16	597	11.7	928	2	086164
17	597	11.7	949	2	09K299
18	595	11.7	1016	2	084880
19	587	11.5	947	2	09JSE2

20	586.5	11.5	914	2	086163	086163 chlamydia p
21	586.5	11.5	928	2	09RB65	09RB65 chlamydia p
22	586	11.5	947	2	092813	092813 chlamydia p
23	585.5	11.5	928	2	092398	092398 chlamydia p
24	583	11.5	978	2	092895	092895 chlamydia p
25	583	11.5	978	2	09RB63	09RB63 chlamydia p
26	575	11.3	973	2	092896	092896 chlamydia p
27	575	11.3	995	2	09K2A1	09K2A1 chlamydia p
28	570.5	11.2	922	2	092965	092965 chlamydia p
29	569.5	11.2	922	2	09K1Y9	09K1Y9 chlamydia p
30	568.5	11.2	922	2	0924H9	0924H9 chlamydia p
31	567	11.1	1276	2	09JRM2	09JRM2 chlamydia p
32	564	11.1	983	2	09PL44	09PL44 chlamydia p
33	562	11.0	1407	2	092899	092899 chlamydia p
34	546.5	10.7	986	2	09PJY3	09PJY3 chlamydia p
35	544.5	10.7	847	2	P71132	P71132 chlamydia p
36	531	10.4	839	2	P77792	P77792 chlamydia p
37	530.5	10.4	772	2	09RB71	09RB71 chlamydia p
38	530.5	10.4	846	2	P71133	P71133 chlamydia p
39	525.5	10.3	841	2	0923A1	0923A1 chlamydia p
40	519.5	10.2	975	2	084417	084417 chlamydia p
41	517	10.2	1723	2	092812	092812 chlamydia p
42	517	10.2	1723	2	09RB59	09RB59 chlamydia p
43	517	10.2	1732	2	09K2C1	09K2C1 chlamydia p
44	512	10.1	392	2	092881	092881 chlamydia p
45	474.5	9.3	1609	2	09RB58	09RB58 chlamydia p

ALIGNMENTS

RESULT 1

AC 09PL47 PRELIMINARY: PRT: 976 AA.

ID 09PL47:

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)

DE POLYMORPHIC MEMBRANE PROTEIN E/F FAMILY.

GN TC0261.

OS Chlamydia muridarum.

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI_TaxId=83560;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MOPN / NIGG;

RX MEDLINE=20150255; PubMed=10684935;

RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,

RA White O., Hickey E.K., Peterson J., Utterback T., Berry K.,

RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,

RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,

RA Eisen J., Fraser C.M.;

RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia

RT pneumoniae AR39.";

RL Nucleic Acids Res. 28:1397-1406(2000).

RN [2]

RP SEQUENCE FROM N.A.

RA Read T.D., Brunham R., Shen C., Gill S.R., Heidelberg J.F., White O.,

RA Hickey E.K., Peterson J., Utterback T., Berry K.,

RA Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,

RA Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G.,

RA Salzberg S.L., Eisen J., Fraser C.M.;

RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.

DR EMBL: AE002293; AAF39130.1;

DR TIGR: TC0261;

SO SEQUENCE 976 AA; 105964 MW; DBEA56158E3CEEA9 CRC64;

Query Match 71.7%; Score 3648.5; DB 2; Length 976;
Best Local Similarity 70.5%; Pred. No. 9.5e-241;
Matches 689; Conservative 121; Mismatches 154; Indels 13; Gaps 3;

QY 1 MKKAFEEFLIGNSLGARLREVPSPRIFLMPNSVDPFTKESLSKISLTJGDTNHTNCTIYDN 60

DB 336 DNGDIFNNNN-----TASKALNPPYRNAIH-STPMNLIQIGARPGYRVLFYDPIEH 386
QY 413 S-NAGVSVEFNKEADOTGVSFSGATVNSADFHQRNLOTKTPAPLTLNSGFLICIEHQAOL 471
DB 387 ELPSSEPLIFNEETGTGTVLSEGEHQNFTDENNFYSRLNTSELRGVLAIVEDGAGL 446
QY 472 TVNRFOTGTGVSALNGVALSCYKKNAGN-----SASNAISITLKHILGNLSLTKSG 523
DB 447 ACYKFFRGRTLLGGAGVITTT-----AGTIPPTSPPTTGVSTTLNHLAIDPLSLTQ 502
QY 524 AEIPLWVEPTNNNYTADTATFSLSDVKLSLIDDYGNSPYESTDLTHALSSQPMESI 583
DB 503 AQAPKIMVPTGTSTYEDSNPTITISGT-LTLRNSNNEPDYDLSLHSEKVLVI 561
QY 584 SEASDNLASDDMDFSGLNV-PIHYGOGIWTMGAKTODPEPASSATTIDPOK---ANRF 639
DB 562 VVVAQKINSQSLDLSTLNGEHYGYGIMSTYVET-----TTITNPTSLGANTR 613
QY 640 HRTLLTLWLPAGYVPSPKHRSPLIANTLWGNMLATFSLKNSAELTPSDHPFGITGG-- 697
DB 614 HKLLIYANMSPFLGTRPERGEREITNALMOSATLALAGLSLSS-----WDEKGA 665
QY 698 ---GLGMVYDDPRENHPFHNRSSGYSAGMIA--GQTHPSLKSQTYTKLNERVAKN 751
DB 666 ASLOGIGLLVHOKDKNGFGRSHMTGYSATTEATSSQSPNFSLGFAPQFSAKKEHESON 725
QY 752 NWSKNY---SCOGEMLSLQSGFLTLKLVGL-YSYGDHNCHEFTYOGENTLSOGTFRSQ 807
DB 726 STSHHFFSGMCIENTLF--KEWIRLS--VSLAYMFTSEHTHYOGLEGNSOGSFFHNH 781
QY 808 TMGGAFFDLPKPKRGSTHILTAFLGALGYSSHTEVGAYPSPSTKPLINVLVP 867
DB 782 TLGALISCYFLQPHGES--LOIYPTTALAIRGNLAFOESGDHAEFSLHRLTDVSLP 840
QY 868 IGVKGSFNATORPQAMVELAYOPVLYROEPGIATOLLASKGIMFGSSPSSRIHMSYK 927
DB 841 VGRBASMKHNRHRYLWLEISYRSLYRQDPELSKLLISOGTWTQATPYNNALGIK 900
QY 928 ISOOTQPLSMVLHFOYHGFYSSTFCNLTNGEIALRF 965
DB 901 VKNTMQVFPKVTLLSDYSADISSSTLSHLYNVAARMRF 938
RESULT 3
QYK2A5 PRELIMINARY: PRT: 952 AA.
AC QYK2A5;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE POLYMORPHIC MEMBRANE PROTEIN E/F FAMILY.
GN CP0285.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Umayam L.A., Uterback T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J.,
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.,
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406 (2000).
DR EMBL: AE002190; AAF38142.1; -
DR TIGR: CP0285; -
SQ SEQUENCE 952 AA: 104196 MW: FICB050DA6B74B5A CRC64;

Query Match

20.9%; Score 1060.5; DB 2; Length 952;

Best Local Similarity 30.6%; Pred. No. 4,4e-64;
Matches 303; Conservative 164; Mismatches 447; Indels 77; Gaps 28;
QY 6 FFELIGNSLSGAREVPSRIFLMPNSVPD-----PTKESLSNK-----ISL 46
DB 8 FLEFLCNFTACMGMTPAYVSLQTDLSLEKFALEDERFPTSPLLDLSLTGLGFPITTF 67
QY 47 TQDTNULT-NCYLDNLRILALILQKTPNGCAVTTIDYLSFDTQKEGTFPAKMLTPESG 105
DB 68 VGNRHNSQDIYLSNKSIDNILLMTSAGAVASCNPL--LSNEDHAFESKNAIGTG 125
QY 106 GAIGVAPSPSPVEIRDTPVIFENNCCRPPTSSNPAANAKIREGAIHAQ--NLYIN 164
DB 126 GALACOG-----ACTTKNKGPLIFESNRGLN-----NASIGETRGALACNGDFTIS 174
QY 165 HNHVVGFKNKFSYVRGAIATANFEVSENOCSFLMDNICIOTNTAGKGAIVAGTSN 224
DB 175 ONOGTFEYFNNSVNMGGALSTNGHCRIOSENAPLLEFN-----NTAPSGGALRENT 228
QY 225 SPFSNNCDLFTTNACCAGATFSP--CSLTGKRGIVYNNRCRKNVETASSEASDGA 283
DB 229 TISDNTRPYIFKNNCGNCGAIDQTSVVAIKNNSGVIFNNNTALSG--SINGNSGGA 286
QY 284 IKVTTLDVTCNGRIFSDNITKNYGAIYAPVYTLVONGPTEYFINNTANKKGAIVYD 343
DB 287 I-YTNLSIDDNPGTILFNNNNCIRDGCAICTOFLTKNSHYFTNN--QGMNGGALMLL 344
QY 344 GTSNKRISADRHAIIFENIVNTNANGTSTANPRRNATTVASSGELILGAGSSON 403
DB 345 QDSTCLLFAEQGIAFQNNNEVLTTC-----RNALHCHTPNS--NLQGANAKVT 393
QY 404 LIFPDTEVSNAGVS--VSNKEADOTGVSFSGATVNSADFHQRNLOTKTPAPLTLNSGF 462
DB 394 TAEFDPIEHQPTTNPLINPNANMOGITLFSAYIPRESDENNEISSKNTSELARNV 453
QY 463 ICIEDHAQTLVNFETOTGVSLSGVALSCYKKNAGNSAS--NASITLHGIGNSLTK 521
DB 454 LSTEDRAGNOFKFTQKGGILKGAHASIATTANSETPSTVSGOYIINNLINLPSILA 513
QY 522 SGAEIPLWVEPTNNNYTADTATFSLSDVKLSLIDDYGNSPYESTDLTHALSSQPM 581
DB 514 KG-KAPTLWIRPLOSAPPTEDNNPTITLSG--VLTLEENRDPYDIDLSLPLQNIHL 571
QY 582 SISEASDNLRSDDMFSGLN--VPHYGOGIWTMGAKTODPEPASSATTIDPOKANRF 640
DB 572 SLSDVTARHINTDNFHPESLATEHYGYGIMSPYVET--ITTNNASI---ETANTLY 626
QY 641 RTLLTLWLPAGYVPSPKHRSPLIANTLWGNMLIATFSLKNSAELTPSD--HPFGITGG 698
DB 627 RALYANMTPLGKRVNFEYOGDLATTPLMOSFTMTSLNSYNRTGSDIERPLETIOGIA 686
QY 699 LGMVYDDPRENHPFHNRSSGYS--AGMIAGQTHFSLSFSQTYTKLNERVAKNNVSK 756
DB 687 DGIIFYHONSIPGAPGRIGTSGYSLQASSETSLHCKISLGFAPQFRTKEISSNNVSAH 746
QY 757 N--YSCOGEMLSLQSGFLTLKLVGLYSYGDHNCHEFTYOGENTLSOGTFRSQTMGAVF 814
DB 747 NTVSSLYVELPW--FOEAF--ATSTVLAYGQDHLHLSHPSHOE--QAEQCYSHTLAAAG 803
QY 815 FDLPMKPFESTHILTAFLGALGYSSSHFEVGAYPSPSTKPLINVLWLPVIGKGSF 874
DB 804 CSFPMQOKSYLHL--SPYQALAIRSHQTAFAEIGNPKPFYSQAKFYMLTLPLIGIGW 861
QY 875 MNATORPQAMVELAYOPVLYROEPGIATOLLASKGIMFGSSPSSRIHMSYKISOOTOP 934
DB 862 QSKFHVPTFEMTLELSYOPVLYQOONPOIGVTLASGSMIDLGNVRYNALGKVVHNOGAL 921
QY 935 LSWLTLHFOYHGFYSSTFCNLTNGEIALRF 965
DB 922 FRSLDLFDLYOGVSSTSTHHLQAGSTLKF 952
RESULT 4


```

Db 359 HCTNS-NLGGANKGYTATFAFDPIEHQHPPTTNPLIFPNMNHQHTILFSSAYIPEDADY 417
Oy 445 QRLNQTAPPLTSLNGLCTIEDHAQLTVNRETGTGVYSLGNGAVLSQYKNGAGNSAS- 503
Db 418 ENNFSSSKNTSELNGLSIEDRAGMOFYKFTOKGILKLGHAASIAATTANSEPTSTSV 477
Oy 504 NASITLKIGLNLSSILKSGAELPLWEPTNNSNNYADTAAPFSLSDVKSLTDGNG 563
Db 478 GSOYIINNLAINLPSILKKG-KAPLMIRPLOSSAPFTEDNNPTTLTSG-PLTLUNEENR 535
Oy 564 SPYESTDLTHALSSOPMLISEASDNOLRSDMDPFGSLN-VPHYGMOGLMTGMNAKTODP 622
Db 536 DPYSDIDSEPLQNIHLISLSDVTAHINTDNHPESLNATEHYGYGIMSYWET--I 593
Oy 623 EPASATITDPOKANRFRHTLLTWLPAGYVPSPKHRSPLANTLGMNMLATESLKMSA 682
Db 594 TTNNAASI--ETANTLRALYAMNTPLGYKVNPEYOGDLATTPLOSFHTMFLSLRSYN 650
Oy 683 ELTPSD--HIFMGITGGGLGMVYODPRENHGPFHMRSSGYS--AGMLAGOTHTFSLKFS 738
Db 651 RTGSDIRPPLLEIOLGADGLFVHONSTIPGAPFRIOSTGYSLOASSETSLHOKISLQFA 710
Oy 739 QTYTKLNERAKNNVSKN--YSCGEMLFSLQEGFLTKLGLYSYGDHNCIHFEYTOGE 796
Db 711 QFFTRKEIGSSNNVSAINTVSLVELPW-FQEAFA-ATSTYLAAGYGDHLSHLPBHQ 768
Oy 797 NLTSQGTFRSOTMGAVFEDLPMPKPFGSTHILTAFLCALGYSSLSHTEVAYPRFS 856
Db 769 E-QAEGTCYSHTLAAIGCSFPWOKSYLHL--SPFYOAIAIRSHQTAFFEEIGDPRKFV 825
Oy 857 TKTPLINLVLPYGVGSMNATQRPQAVTELAVQPVLYXROPGLATQLLASKGIMFSG 916
Db 826 SOKPRYNLPLPLGLOGKMQSKFHVPTMTLELSTVQPVLYQOPQICVTLGLASGSGMDILG 885
Oy 917 SPSSHHASYSKISOQTOPLSWLTLFHGYHGFYSSSTFCNYLNGEIALRF 965
Db 886 HNYRNALGYKVHQTALFRSLDLFLDYOGSSVSSSTHHLQAGSTLKF 934

RESULT 6
OyRB60
ID 09RB60 PRELIMINARY: PRT: 946 AA.
AC 09RB60:
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN E/F FAMILY (POLYMORPHIC MEMBRANE
DE PROTEIN E/F FAMILY).
GN PMP_18 OR CP0283.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-J138:
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hatori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AR39:
RC MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA Bowman C., Dodson R., Quinn M., Nelson W., Deboy R., Kolonay J.,
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).

```

```

DR EMBL: AP002546; BAA98677.1;
DR EMBL: AE002189; AAF38141.1;
DR TIGR: CP0283;
SO SEQUENCE 946 AA; 103611 MW; DCBF6B3D04AC452C CRC64;

Query Match 18.7%; Score 951.5; DB 2; Length 946;
Best Local Similarity 29.1%; Pred. No. 1.2e-56;
Matches 281; Conservative 168; Mismatches 416; Indels 101; Gaps 34;

Oy 35 PTKESLSKKISLTGDTNMLT-----NCYLDNLRYLALILQKTPNEGAAYTTIDYL 84
Db 47 PLIDITLN---MTPYSHRATLFGVRDITNODIVLDHQSISWFENFSDGALSCSK-L 102
Oy 85 SFEDTQKEGIFYFAKMLTEGSGAIGYASPNSEYVIRITQIPVIFENNCCRPFTSSAPN 144
Db 103 ALTNT-KNQLIFLNSFAIKRAGAM-YVNGN--FDLSNHQSIIFSGNL-----SFPN 150
Oy 145 AA--VNKIREGGAHMOULYINHNHDVYGFKKNFYVGAISTANTFVSENOGSCFLFM 202
Db 151 ASNFADTCTGCAVLCSSKVTTSKNOGTAFVINKAKSSGALQAAIINIKONTGCLFF- 209
Oy 203 DNICIQTNAG--KGAIVAGTSNFSNNDLFFINNACCAGAI-FSPICSLTGNRGN 259
Db 210 -----NNMAGGTAGALFANACR-IENNSQPIYFLNNGSLGAIIRVHOCILTKNGS 262
Oy 260 IVFYNNRCFKNVETASSASDGAIKVTRLDVTGNRGRIFPSDNTKNYGAIYAPVYT 319
Db 263 VIFNNN--FAMEADISAHNSGCAIYCS-CSIKDNPGLAFDNTARDAICLCTOSLT 319
Oy 320 LYDNGPTFEINNIANNKGAIIYIDGTSNKSISADHAIIFENNTVNTNNGTSTANP 379
Db 320 IODSGPVYFTNN-QCTMGCAIWLRODGACTIONLIFNNRHKOTFSN--HVSVC 376
Oy 380 PRNATVYASSGCELLGAGSSONLIFYDPI--EVSNAQSVSFKNEADQTSVFSVG 435
Db 377 TRNVSLTYGASGQ-----HSATFYDPIQRTIONS--IQFNNPHELGLIFLSS 425
Oy 436 A-----TNSADF--HQRNLOTKTPAPLTLNGLCITDHAQLTVNRETGTGVYSLGNG 488
Db 426 AIIPTSTSRDEFISHFRN-----HIGLYNGTLALDEKAWKYKFDQCGTIRLCSR 478
Oy 489 AVLSCYKNGAGSASNASITLKLGLNLSILKSGAELPLWEPTNNSNYTADTATF 548
Db 479 AVFSTTDEQSSSVGSVINNNLAINLPSIL-GRRVAPKAIPTGTSAPYSEDNPII 537
Oy 549 SLSDVKLSLIDYGNPSYESTDLTHALSSOPMLISEASDNOLRSDMDPFGSLN-VPHYG 607
Db 538 NLSG-PLSLDDENDLPDVTADLAPIDAEVPLVLLDVTAKHINTDNFPEGLNTQHYG 596
Oy 608 WGLTWGMAKTODEPASSATITDPOKANRFRHTLLTWLPAGYVPSPKHRSPLIANTL 667
Db 597 YGVGWSPIWIEITTSDSIS-----EDTVNLTLRQLYGDMPTGYKVPENKGDIALSAF 651
Oy 668 W----GNMLLATESLKNASDELTPSDHPFGITGGGLGMVYODPRENHGPFHMRSSGYSAG 724
Db 652 WOSFHNLFATLRYOTOQOQIAPT-----ASGEATRLVHONSNDAAGFHEATGYSLG 705
Oy 725 MA--GQHTFSLKFSQTYTKLNERAKNNVSKNYSCGEMLFSLQEGFLTLTVLY 781
Db 706 TTSNTASNSHSEGVNFSOLFNSLYESHSDNSVASHTTYALQIINNWLDERSTASLA-Y 764
Oy 782 SYGDHCHHFFYTQGEN--LTSQGTFRSOTMGAVFEDLPMPKPFGSTHILTAFLCALGY 839
Db 765 SVSN---HHIKASGYSGKIOTEGCKCYSTLGAALSCSLQ-WRSRPLHPTFIDAIYVR 820
Oy 840 SLSLHTEVGAAPRFSSTKPLINLVLPYGVGSMNATQRPQAVTELAVQPVLYXREP 899
Db 821 NQOTAFQESGDARFVHKPLYNLTPLGLOSARESKFRPLTYVNIETLAVQPVLYQOP 880
Oy 900 GIATOLLASKGIMFGSGSPSSHHASYSKISOQTOPLSWLTLFHGYHGFYSSSTFCNYLNG 959
Db 881 EVNVSLESSESSGMLSLSTLARNALAFKGRNQIIFPRLSVFLDYOGSSVSSSTTHHYHA 940

```

```
Db 80 LSTNSSGAFV--ESMTSETAIDNADLFECNNVCTHOG--GGAINATGLISFKNQNIL 135
OY 128 IFENNTCCARFPSSNPMAVNKIREGGAHAONLYINHHDDVVGPKNRSYVRCGAIISA 187
Db 136 FNNNTTIGTFOPVALTERNR--GGALYSSIELINHS-LNFNNNSGOMGAVSTI 191
OY 188 NTFVSENOCEFLFMDNICIQ--TNT---AGKGAIVAGTSNSESNNCDLFEFFINACC 241
Db 192 ONLVYIKNTSGIYAFENHHTDHPNFATILARGAVGCGACELSHNTGPVFNNSVYG 251
OY 242 AGCAIIFS-PICSLTGNRNIYVYNNRCFKNVETASSEASDGAIKVYTRLDVTCNCRIF 300
Db 252 YGCAISTGQOCIFRDKDKLFIINNSALGMHNT--SAQNGAVISAGEFGLNNKGPY 309
OY 301 FSDNITKNGCAIYAPVTVLVNDGPTVFINNIANKGAIYIDGT-SNKKISADRAILIF 359
Db 310 FENNMASTYIAGAIACNNLNFQENGPYFLNNSA-LYGGAFHLFASPAANYIHTGSDILF 368
OY 360 NENITVNTNAN-----GTSTANPPRRNATTVASSGELLGSSQNLIFEDP 409
Db 369 NNNTLSTTGMSAGLRKLFYIPGTNN-----NPITLS-----LGAKKOTRIYFDL 415
OY 410 IE-----VSNAVSVSEFKKADQTSVFGATVNSADFHORNQOTKT-----PA 454
Db 416 FQWGLKKNATPPENSPHTVITNPDSFGAVVFSYKNI-SSDLQAHIMIASKTHNQIKDS 474
OY 455 PTLISNGFLCTIDHAQ-----LTVNRTOTQGVVSLCNCGLAVSCYKKGACNSASNAI 507
Db 475 PPTLFGTMSIENGAEFEFFNGPLT---QESTSLALOODSLITLVGK-----DASL 522
OY 508 TLKHGLNLNLSLTK-SGAIFPLWVEP---TNNSNYTDATATSLSDVKL-----SL 557
Db 523 TITHGLIILPGLLNOGTTAPRIARVNPDMOTONTNSNOA-PVSTEVAIYOKLFFSGSL 581
OY 558 IDDYGNSPYESTDLTHALSSQPMLSISEASDQNRSD--DMDFGLNVPYHGMOLMTW 614
Db 582 VDENEYSYDSCDLSCRKANOPILHETIETDQSLSDMKNTLNTSLPHYGYGLWLS 641
OY 615 GWAKTQDPBPASSATITDPOKANR-----638
Db 642 NMHTT--TRTVSLJNSTEOTJANNNSIOEOKNTSETFDSNSTTAKLPSIRASTGCTPLA 699
OY 639 -----FHRTLLTWLPAGVYSPGKHSPLIANTLWGNMLATESKNSAEITPSDHPW 692
Db 700 TTDVTVTRISLAVSWPTIGYIADPARCDLIANNLVSSCRNTTLVLR--LLPDD-SWF 755
OY 693 GTTGGGLGMVYQDPRENNHGFHMSSGY--SAGMIAGOTHTFSLFSQTYTKLNERVAK 750
Db 756 ALQGSAAITLFTKQKRRLDYHGYSASAKYAISSQASGAIGHKFLFSQSDTMMKERTN 815
OY 751 NNVSCKNSCOCEMLFSIOEGFLTKLVG--LYSGDNHCHAFPTOGENULTSCGTFRSOT 808
Db 816 NMISSRYYL--SALCFECPMEDRIALLIGAAYVNGTHTYNYV--CTKKFKGCHNST 870
OY 809 MCGAATFFDL-PKMPGSGTIIITAPFALGALGIVSSLSHFTEVAVAPRPSFTKPLIVLVP 867
Db 871 LGGSLRCELRDMPFOS--IMLTPICALISRTPEASIOEGDLARIFFLCKPPIHAYVSP 928
OY 868 IGVKGSFNATQRPQAWTVELAYQVLYRQEPGIATOLLASKGIFWGSGSPSSRHAMSYK 927
Db 929 IGIKGVYSSNMKPJVSCEMEVAYOPTLWYKRPILNTVLKKNNGSMETTNTPLAKHSF-YG 987
OY 928 ISOOTQPLSLMLHFQYHGFYSSSTFCANLNGEIALRF 965
Db 988 RGSLSLKFSYLKLFANYQAOVATSTVSHYMNAGALVF 1025
```

```
RESULT 9
ID 084879 PRELIMINARY: PRT: 1013 AA.
AC 084879:
DT 01-NOV-1998 (TREMblrel. 08, Created)
01-NOV-1998 (TREMblrel. 08, Last sequence update)
```

```
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE PUTATIVE OUTER MEMBRANE PROTEIN G.
GN PMPC.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxId=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RX MEDLINE=9900809; PubMed=9784136;
RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
DR EMBL: AB001360; AAC68469.1;
SQ SEQUENCE 1013 AA; 107366 MW; F0927743C0A651DD CRC64;

Query Match 12.9%; Score 655.5; DB 2; Length 1013;
Best Local Similarity 25.7%; Pred. No. 2.1e-36;
Matches 261; Conservative 152; Mismatches 388; Indels 213; Gaps 43;

OY 39 SLNKRISLTGDTNHLNCLYDLNLRILATLOKTPNDEGAVTITDYLSPFDQKEGITYAK 98
Db 84 NIIAGSTVYIGRCHSLT--FENIR-----TSTNGAALS-----DSANGLF--- 121
OY 99 NITPESGCAIGYASNPSPVEIRDTIGPVIFENNCCRPFTSSNP-NAV----- 147
Db 122 --TIEGFKLSLSSNONSLLAVL-----PATTNNSQPTTSTSTNSGITYKTDLLLLL 174
OY 148 -----NKIR-EGGAHAONLYINHHDDVVGPKNRSYVRCGAIISANTFVSENO SC 198
Db 175 NEKFSFYSNLVSGDGAIDAKSLTVOGISKLCVFOENFAQADGACQVVTSPSAMANEP 234
OY 199 FLEMONICQNTAG-KGAIIYA-----GTSNSE-----SNDCDLFINNACCAG 243
Db 235 IATIANV-----AGVRGGIAAVODGCGGVSSSTEDPVVFSFRNRAVERDGVNARVG 288
OY 244 GAIFSPICSLTGNKGNIVFYNN--RCFKN-----VETASSEASDGA 284
Db 289 GGIYS-----YGNVAFLNNGKTLFLNNVASPYIAEPTNGQASVTSNVDGGA 340
OY 285 KYTRLDVTGN-----RCRIEFSNITKKNYGAIYAPVTVLVNDGPTVFINNIAN 334
Db 341 PCKNGAQAAGSNNSGSVSFDEGVVFFSSNVAAGKAIYAKKLSVANCPOVQGLNIAN 400
OY 335 NKGAIYIDGTNSKISADRHAIITENET---VNVNTANCTISANPPRNATTVASS 390
Db 401 D-GGAIYIGSESLISADYGDIFDGNLKRakeNAADVNCVYSS--QALSKSGC 454
OY 391 SCEILLGAGSSONLIFYPIIEVSN-----AGVSYFKEADQTSVYFGATVNSADF 443
Db 455 GKITTLRAAKHQIILFNPDIEAMANNQPOSSSEPLKINDGCGYGDIVF--ANNSITLY 512
OY 444 HORNILOTKPPAPLITISNGFLCTIEDHAOLTVNRFOTQGVVSLCNCGLAVSCYKNGAGNS- 501
Db 513 QN-----VTEOGRIVYLRERAKLSVNSLSOTGGSLYMEASITLDFVPPQPOPP 562
OY 502 ASNASITLKHIIGLNLNLSLTKSGAEIPLWVEPTN--NSNNYTAQTAATFSLSDVLS 556
Db 563 AANOLITLNLNLSLSSLLANNA-----VNPPTNPPAODSHPAIIGST--TAGSVTISGP 617
OY 557 LIDDYGNSPYESTD-----LTHALSSQPMLSISEASDQNRSDMDPFGSLNVPYH 606
Db 618 FREDLDDTAYDRYDMLGNSQKIDVLKLDGTOP--SANNPSDLTL-----GNEMKY 667
OY 607 GWQGLTWGMMAKTQDPBPASSATITDPOKANRPHRTLLTWLPAGYVSPKHSPLIANT 666
Db 668 GYGNSWKLAW-----DPTANNNGPYTLKATWTKTGYNGPPEVAVSLPNS 712
OY 667 LMGNNMLATESLKNSEAEITPSDHPW-GITGGGLGMVYQDPRENNHGFHMSSSYSGM 725
```

RA Bowman C., Dodson R., Gwinn M., Nelson W., Deboy R., Kolonay J.,
RA McClarty G., Salberg S.L., Eisen J., Fraser C.M.,
RT Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39.
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL: AP002546: BAA98653.1;
DR EMBL: AE002193: AAF38165.1;
DR TIGR: CP0308;
SO SEQUENCE 936 AA: 100105 MW: 3981DB3C950AF95A CRC64;

Query Match 12.7%; Score 644; DB 2; Length 936;
Best Local Similarity 27.0%; Pred. No. 1,2e+35;
Matches 262; Conservative 146; Mismatches 392; Indels 172; Gaps 44;

OY 75 GAAVTTTDLSEFDKQEGIVFAKNTPESSGAIGY-----ASPSPTVEIR 121
DB 54 GTTYSLSLDVSTFQNGALGIPLAGCGFLBAGDLPFGQNHAKLFAFINAGSAGTAVST 113
OY 122 DTGCVIIFENN-----TCCRPFSSNNMAVANKIREGALHAOMLYINHHHDVYGFMK 174
DB 114 SAADKNLLFNDPSRLSIISCPFLSLPTGQCALKSV-----GNLSLTGNSQII-FTQ 164
OY 175 NFSYVGCAISTANFVFVSENOGCLFMDNICTQRTACKGCAIYA-GTNSPESNCDL 233
DB 165 NFSNDGNGVITNKN-FLLSGTQSFASFNRN---QATGKGGGVYATGTTTENSFGIYS 220
OY 234 FFINNACCAGAIIFS-PICSLTGNRCNIVFYNNRCKENVETASSSEASDGAIKVTT--R 289
DB 221 FSONLAKGCGALYSTDCNSITDN-FQVI FDSGSAWEAAO-----ACGCACTTTDKT 273
OY 290 LDVTGRCGIFESDNTKRYGCAIYAPVYTLVDCNGPTFYFINNIANK-----GCATYIDG 344
DB 274 YLTGTGK-NESTNNMTALTYGAIISGLKVISAGGPTLFQSNISGSGAGCGGAINIAS 332
OY 345 TSNKSIADRAHAIIFENIVTNTNANGSTSNANPRRAITVASSSGEILGAGSSONL 404
DB 333 ADELALISATSGDITFNNQVTN-----GSTST-----RAHINIDRAKTSIRATGOSI 382
OY 405 IFYDPIEVSNAVSYS-----FNKEADQTSVSESGATVNSAD-FHORNLOTKTP 453
DB 383 YEYDPI--TNPGTAASTDTLNLADANSEIEYGCAIVSEKLSPTFEKAIANVTSTIR 440
OY 454 ABLTISNGCLCEDHAQLTVNRFOTGCVVSLGNGAVLSICYNKAGNSASMITLKHIG 513
DB 441 GPAYVLARGVLVLDGVTVEFKDLTOSPGSRILMDG-----GTTLSAKEANLSINGLA 492
OY 514 LNLSSILKSGAEIPLLVPEPTNNSNNYTDATATFSLSDVKLSLIDDYNSPESDILFH 573
DB 493 VNLSSL-----DGTNKAALTEADKNISLSGT-IALLIDTEG-SFEENINLMS 538
OY 574 ALSOPMLISEASDN-QLRSDMDSEGLNVP--HYGMOGLTWGNAKTODEPASSATI 630
DB 539 A-STYPLELELTAGANGTTLTALSTLTQEPETHGYGQNGNOLSMAN-----ATSSKI 591
OY 631 TDPKANRHRKTLTLTWPAGYVPSPKHRSPLANTLKGNNMLATESLKSASELTPSDHP 690
DB 592 GS-----INMTRTGYIPSPERKSNPLNLSLWGN-FIDIRISINQLIETKSSGEP 638
OY 691 F-----WGITGGGLGMVYODPRENHGCFHMRSSGYAGAIACOTHTFSLKFSOTYKLE 746
DB 639 FERELM--LSGIANFFYDSMPTRHGFRHISGVALGTTATTPADQLTFA--FQQLA 693
OY 747 RYAKNNVSSKN-----YSCQGEMLFSLQEGFL-----LTKLVGL---- 780
DB 694 R-DRNHITTKNGDRTYGASLYFHHTEGLFDI--ANFLMGKATRAPWVLSLSQIILPSFDA 751
OY 781 ---YSGDHNCNHFFYQGNLTSOGTFRSOTMOGAFFDLPKMPSTHIL--TAFPLA 835
DB 752 KRSYLTLDHNMKTYT--DNSIITKGSWRNDACADAGASLPF-VISVPYLKEVEFEVAV 808
OY 836 LCIYSSLSHFTVGAVPRFSFTKTPLINLVPIGVKGSFNNATQRQAWTVELAYOPV- 894

DB 809 QYIYAHQODFYERYAGRAFN-KSELINVEIPIGVT---FERDSKSEKTYDULITMIL 864
OY 895 -YRQEPGINTOLLASKGIFGSGSPSSRIHMSYKISQOTPLSWTLHFQYHGFYSSSTF 953
DB 865 AYRRNPKCOTSLIASDANMAYGTNLAROGFSYRANHFOVNPHEIFGOF-AFEVRSS 923
OY 954 CYNUNGEIARF 965
DB 924 RNT-NTNLGSKF 934

RESULT 12
P71135 PRELIMINARY; PRT: 926 AA.

AC P71135;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE PUTATIVE 98 KDA OUTER MEMBRANE PROTEIN.
OS Chlamydia psittaci.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83354;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ABORTION S26/3;
RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
RL Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL: U72499; A818188.1;
SO SEQUENCE 926 AA: 98439 MW: 3E75E52F594750F CRC64;

Query Match 12.5%; Score 636; DB 2; Length 926;
Best Local Similarity 24.0%; Pred. No. 4e-35;
Matches 241; Conservative 157; Mismatches 365; Indels 240; Gaps 41;

OY 41 SNKISLTGTHNL--TNCYLDNLRYILATLOKTPNCGA-----AVTIDY--L 84
DB 82 ADNLTPKGNHSLITNA-----NAGAPACIANTADKILTLIDFSKL 125
OY 85 SFEDTQKEGIVFAKNTPESSGAIGYASPSPTVEIRDTIGVIFENNCCRPFTSSNPN 144
DB 126 SFPEKSSLVNTHGK-AMKSGALNANNAS-----ILPDQN-----YSKEN-- 166
OY 145 AAVNKIREGALHAOMLYINHHHDVYGFKNFSYRGCAISTANFVSENOGCLFMDN 204
DB 167 -----GGAISCKAFSLTSGSKSEISFTNSTAKKGAIATGIAHLSDNOGITRFSGN 218
OY 205 ICIOTACKGCAIYAGTNSPESNCDLFTINMCC-----AGCAIFPISLTGNRCN 259
DB 219 TAVNS-----GCAVVSSEASMTIAGNN-HVAFSNNAVSGSSDCCGAIH--CSKTGSAPT 269
OY 260 IVEYNNRCKENVETASSSEASDGAIKVTTRLDVTGNRGRIFESDNTKRYGCAIYAPVY 319
DB 270 LTIROKKV-----LIEENISSAKGAIITDKI 298
OY 320 LYDNGPTFTNN--IANKGCAIYIDGTSNKSISADRAHAIIFENIVTNTNANGTST 376
DB 299 LTSGGPTAFINNKVTHATPKGAIIGIANGECSLTAEHDITFDNNLMATODNAT----- 353
OY 377 ANPRRNATIVASSSEILLGAGSSONLIFDYPIEV-SNAGVSYSFNK-EADOT--GSVY 432
DB 354 ---IKRNAINIEGNCGFVLRABASGKTTFSPYDTIVEGNAADLTLNKAEGDKTYNGRII 410
OY 433 FSGATVNS-----ADFHORNLOTKTPAPLTLSNGFLCIEDHAQLTVNRFOTGCVVSLGN 487
DB 411 FSGEKLTQGAADVAD---NLKTTFTQPTTLAAGELVLSGVEEAKTYVVGAGSLIMLD 466
OY 488 GAVLSCTYKNGAGNSASNASITLKHIGLNLST-----LKSAGELPLLM-----VEP 533
DB 467 A-----GTRLASATEDATLTNLAINPNTLDGKKFAVAVDAVAGKNVTLGSAIGVDP 518
OY 534 TNN--SNNTADTAATFSLSDVKLSLIDDYGNSPESDNL-THALSSQMLISEASDNO 590

Db 341 -----TTGSTDTPKRNAINIGSNGKFTLRAAKNHTIFFEYDPTSEGTSSDVLKINNGSA 395
QY 426 -----DQTGSVVFSSGATVNSADPH-QRNLOTKTPAPLITLNSGCLIEDHAQLITVNFQ-T 479
Db 336 GALNDYQGTILFSGEITLADDELKVNADNLSKSTFQVSLSGKLLLOKGVTLTESTSFSQNA 455
QY 480 GGVNSLGNCAVLSCKYKNGASNASITLKHIGLNLSSILKSGAEIPLLWEPINNSNN 539
Db 456 GSLGMDSGTTL-----STAGSITITMLGINVDSL--GLKQY-----S 493
QY 540 YTAOTAAFTSLSDVKLSLIDYGNSPYESTDLTHALSSQPMLSISEASDNLQSDMDFS 599
Db 494 LTAKGASNKVIVSGKLNLDIEGNI-YES---HMFSDQLFSLKITYDADVDVNDIS 548
QY 600 GL-NYP-----HYGMOGLMTWGMMAKTODPEPSSATITDPOKANKRHTLLTLPLAGY 652
Db 549 SLIPVPAEDPNSEYGFQGMVNW-----TTDTATNTEAT-----ATWTKTGF 592
QY 653 VPSPKHRSPLANTLMGNMLLATESLKNASAE--TPSDHP--FMGITGGGGMVYODPR 708
Db 593 VPSPERKSAVLCNTLMG-VFTDIRSLQOLVEIGATGMEKQGFV---VSSMTNPLHKTGD 648
QY 709 ENHDPFHMRSSGYSAGMLAGQTH-----FSLKFSQTYTKLMERY-AKNN-----752
Db 649 ENRKGFRHTSGY---VIGSAHTPKDLETFPACHLFARBDKCFIAHNSRTYGTLEF 705
QY 753 -----VSSKNVSCQEMFLSLO--EGF-----LTKLVGLVSYGCHNCHHTTQGENLTS 800
Db 706 KHSHTLOPNTLRLGRARFSESATKFPREIPLALDVOVSFSHSDNRMTHTYSLE--S 763
QY 801 QGTFRSQTMGAVFEDLPKPPGSTHILAPFLGALG---IYSSLSHTEVAGYRST 857
Db 764 EGSMSNECIAGIGIDLDP-VLSNHPLEKFTIPQMKVEMVYVSONSFESSDORGSFI 822
QY 858 KTPLINLVLPICGVKSFNNATQPOAWVELA--YQPVLYROEPIATQLASKIWFCS 915
Db 823 GR-LNLSTIPVCAK--FVOG-DIGDSTYDLSGFFVSDVYRNPNPOSTATLVMSPSWKIR 878
QY 916 GSPSSRHA 923
Db 879 GGNLSROA 886

RESULT 17
Q9K299
AC Q9K299 PRELIMINARY; PRT: 949 AA.
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE POLYMORPHIC MEMBRANE PROTEIN G FAMILY.
GN CP0302.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heideberg J.F.,
RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA Bowman C., Dodson R., Guim W., Nelson W., Deboy R., Kolonay J.,
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.,
RT Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39. "
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL: AE002192; AAF38159.1. "
DR TIGR: CP0302. "
SQ SEQUENCE 949 AA: 101357 MW: A00809E16C699BE3 CKC64;

Query Match 11.7%; Score 597; DB 2; Length 949;

Best Local Similarity 26.1%; Pred. No. 1,9e-32;
Matches 253; Conservative 148; Mismatches 361; Indels 206; Gaps 47;
QY 36 TRESLSNKLSTLGD-----THNLNLCY---LDNLRYI-----LALQKTPNNGA 76
Db 66 TPRTSATYTSLTGDFEYEPKGTPLSDSCFKQTDNLTFLNGHSLTFGFIADCTHAGA 125
QY 77 ATTIDYLSFFTOVEGIFAFKNTLPESGAGIYVSPNSPTYEIDTIGPVTFENNCCR 136
Db 126 AASTT-----ANKNLTFSGFSLTSFSSPSPTV---TTGQ-----157
QY 137 PETSNNPNAVNRKIREGAIHAONLYINHHNHDVCFMKFESVRCGAISTANTFVSENO 196
Db 158 -----GTLSSAGVNLNT---KLVYAG--NFSIADGAIKGA-SFLITGTS 199
QY 197 SCFLFMDNICIOTNPAKGAIVACTSNSFESNCDL--FFINACGAGAIFFPICSLT 254
Db 200 GDALFSNN-----SSSTKGAI-ATTAGARIANNNGYVFLSNIASTSGAIDDEGTSIL 253
QY 255 GNRGNIVFYNNGCFKNVERASSEASDGAIKYTRLD---VTGNRGRIFESDNTTKYK 310
Db 254 SNKKFLYFEGN-----AAKTGGAIQNTKASGSPELLISNKKTLIFASNAETSG 303
QY 311 GAIVAPVTVLVNCPTEYF--NNI--ANNKGAIVYDGTNSKISADRAHIIENIYVNV 367
Db 304 GAIIHAKKLALSSGFTFELRNVSATPKGSAISIDASELSAETGNITFVRNTLT-- 361
QY 368 TIANGTSTANPPRRNATVASSSGEILGAGSSONLIFYDPI--EVSNAGVSVSNKEA 425
Db 362 -----TTGSTDTPKRNAINIGSNGKFTLRAAKNHTIFFEYDPTSEGTSSDVLKINNGSA 416
QY 426 -----DQTGSVVFSSGATVNSADPH-QRNLOTKTPAPLITLNSGCLIEDHAQLITVNFQ-T 479
Db 417 GALNDYQGTILFSGEITLADDELKVNADNLSKSTFQVSLSGKLLLOKGVTLTESTSFSQNA 476
QY 480 GGVNSLGNCAVLSCKYKNGASNASITLKHIGLNLSSILKSGAEIPLLWEPINNSNN 539
Db 477 GSLGMDSGTTL-----STAGSITITMLGINVDSL--GLKQY-----S 514
QY 540 YTAOTAAFTSLSDVKLSLIDYGNSPYESTDLTHALSSQPMLSISEASDNLQSDMDFS 599
Db 515 LTAKGASNKVIVSGKLNLDIEGNI-YES---HMFSDQLFSLKITYDADVDVNDIS 569
QY 600 GL-NYP-----HYGMOGLMTWGMMAKTODPEPSSATITDPOKANKRHTLLTLPLAGY 652
Db 570 SLIPVPAEDPNSEYGFQGMVNW-----TTDTATNTEAT-----ATWTKTGF 613
QY 653 VPSPKHRSPLANTLMGNMLLATESLKNASAE--TPSDHP--FMGITGGGGMVYODPR 708
Db 614 VPSPERKSAVLCNTLMG-VFTDIRSLQOLVEIGATGMEKQGFV---VSSMTNPLHKTGD 669
QY 709 ENHDPFHMRSSGYSAGMLAGQTH-----FSLKFSQTYTKLMERY-AKNN-----752
Db 670 ENRKGFRHTSGY---VIGSAHTPKDLETFPACHLFARBDKCFIAHNSRTYGTLEF 726
QY 753 -----VSSKNVSCQEMFLSLO--EGF-----LTKLVGLVSYGCHNCHHTTQGENLTS 800
Db 727 KHSHTLOPNTLRLGRARFSESATKFPREIPLALDVOVSFSHSDNRMTHTYSLE--S 784
QY 801 QGTFRSQTMGAVFEDLPKPPGSTHILAPFLGALG---IYSSLSHTEVAGYRST 857
Db 765 EGSMSNECIAGIGIDLDP-VLSNHPLEKFTIPQMKVEMVYVSONSFESSDORGSFI 843
QY 858 KTPLINLVLPICGVKSFNNATQPOAWVELA--YQPVLYROEPIATQLASKIWFCS 915
Db 844 GR-LNLSTIPVCAK--FVOG-DIGDSTYDLSGFFVSDVYRNPNPOSTATLVMSPSWKIR 899
QY 916 GSPSSRHA 923
Db 900 GGNLSROA 907

RESULT 18

```
Db 189 ANODESHRGYGAIVSISIP-----GSPITFADNOELLFOENEGELGAIIVDOCAITF 242
OY 227 ESNNDLFFINNACGAGAIFFSPICSLTGNRNIIVEYNNRCKNVTASSEASDGAIVY 286
Db 243 ENNFQITFSFNKASFCGAVYSRYCNLSQMGDTLF-----TKNAAKVCGAIH- 293
OY 287 TTRLDVTCGRGRFFESDNTITKNYCAIYAPVYTLVD-NCPTTFINNIA--NNKCGAIYIDG 344
Db 292 ADYVHIRDCKGSIVFEEN--SATAGAIIVAVACDINAOGVPRFINNSALCLNGAIYMOA 350
OY 345 T--SNSKISADRAHAIIFENIVTVN--TNANGSTSANPPRRNITAYASSGCEILLAGSS 401
Db 351 TOSILRLHANOCDIEFCCKNKSQFISHINSTNFTN---NALTIOGAFRESLSANEG 406
OY 402 ONLIFYDPI---EVSNAQSVSFNKEADOTGSVFSGATVNSADFHORNLOTKTP---A 454
Db 407 HRICEFDPIISATENYNS--LYINHQRLLFAGAVIFSGARISPE--HKKNKKKTSIINO 463
OY 455 PLTSLNGFLICEDHAOLTVNRFTOTGCVVSLGCAVLSCTKNGAGSASASITLKHIGL 514
Db 464 PVLCSGVSISIEGAILAVRSFQEGGLALGFGSKLTT---QGNKSEKDKIVITNLGF 519
OY 515 NLSILKSG-AEIPILMVEPTNNSNNTADPTATFSLSDVKL-----SLID--DYGNSP 565
Db 520 NLENLSDPAEI-----RATEKASIEISGVPRVYCHTESEYENHETASKP 565
OY 566 YESTDLTHALSSQPLSISEASDNOLRS---DDMFSGLVNPHYGNOGLWTMGMAKTODP 622
Db 566 YTTSTL--LSAKKLVTAOSRPEKDIONLIASEYMG-----YVGSGSEFMSPDNFK 617
OY 623 EPASSATTIDPOKANFHTLLTLWLPAG--YVSPKHSPLANTLWGMNL-----LA 674
Db 618 E-----KTIILSMPTGTEFSLDPKRRGSEFIPTTLMSFTSGLINASNIY 661
OY 675 TESLKSATLPDSHPFWGTGGCLGMVYODPRENHPRGHMRSSGASACIMAGOTHTES 734
Db 662 NNNYLNNSVEIPLQH--LCVFGPVIQIMQONPKOSSNNILVQHAGNNG--ARIPSEN 717
OY 735 LKFSQTYTKLERNYAKNNVSSKNYSQGEMLF---SLOEGFLTKLVGLYSYGDHN---C 788
Db 718 TILSALTLQFSSSSQONVADKSHA---QILIGTSLNKSQWQALSLRSSFTYEDSGVMK 774
OY 789 HHFTYTGENTLSQGTFRSQTMGCAVEFDLPKPFSGTHILLAPLGAIGIYSSL--SHT 846
Db 775 HVEPYG---TSRGSWRNKGSGVGSVAY--PKGIRYLMKTFEVDLQ--YTKLVNPFV 828
OY 847 EVGAYPRSFSTKPLINLVPIGVKSGFMNAQPOQAM--TVELAYQPVLYROBGIATOL 905
Db 829 ETCYDPRYTS--SEMNNLSLPIGIALEMRIFGRSSSLFLOVSTYIKDLRNPPOSASL 887
OY 906 LASKGIMFGSGSPSSRRAMS-----YKI 928
Db 888 VLNHVTMDIQGVPLGKALNITLNTSTIKYKI 918

RESULT 20
086163 PRELIMINARY: PRT: 914 AA.
AC 086163:
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE OUTER MEMBRANE PROTEIN 5 PRECURSOR (FRAGMENT).
GN OMP5.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI:TaxID=83558;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CDC/CWL-029/VR-1310;
RA Knudsen K., Madsen A.S., Mygind P., Christensen G., Birkelund S.;
RT Identification of two novel genes encoding outer membrane complex
```

```
RT associated surface layer proteins in Chlamydia pneumoniae.
RL Submitted (Aug-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: A0001311; CAA04671.1;
KW SIGNAL.
FT SIGNAL. 1 15 POTENTIAL.
FT CHAIN 16 >914 OUTER MEMBRANE PROTEIN 5.
FT NON_TER 914 914
SO SEQUENCE 914 AA: 95603 MW: 88F33BAB680FF5E3 CRC64:

Query Match
Best local similarity 25.6%; Pred. No. 9, 4e-32;
Matches 264; Conservative 143; Mismatches 377; Indels 249; Gaps 48;

OY 1 MKKAFFFFLIGNSL-----SGLAREV-PSRIF-----LMPNSVDPPTRES 41
Db 1 MKGQFSLVLSLTLACFTSCSVFAATVAENIGPSDFDSTNGTYTPKNTTGIDYTL 60
OY 42 NKISL--TGTHNLT-NCYLDNR-----YLLALO-KTPNECAATIT----- 81
Db 61 GDTTLNLDGSAALTKGCESDITESLFAKGYSLSFLNKKSSAECAALVTTDKNLSLT 120
OY 82 --DYLSEFDTQKEGIYFAKNLTPESGAIGVAPSPNPTVEIRDTIGPVIFENNTCCRP 139
Db 121 GFSLLTFIAPSSVT-----TTPSGKAVKCGDLT-----EDNCTILFKDYC- 165
OY 140 SSNPNAVNRKIREGAIHMONLYINHHVDVGFKNFSYVGAIGAIANTFVVENOSCF 199
Db 166 -----EENGAIISPMKLSLKSTGSISEPKKS----- 193
OY 200 LEMDNICLOTNTAKGCAIYAGTSNFSFNCDLFTINN--ACCAGATFSP--ICSLTGR 257
Db 194 -----SATGKKGALCATGTVDITNTAPTLSSNNIAEAGAIINSTGCTITGNT 244
OY 258 GNIVFYNNRCKNVEITASSEASDGAIAKVTTRLDVTCGRGRIFESDNTITKNYCAIYAPV 317
Db 245 -SLVFSN-----SVTATAGNGALSGADVYITIGNOS--VFFSGOAVANGCAIYAAK 295
OY 318 VTLVD--NGPTTFINNIA--NKGAIYIDGTSNKSISADRAHAIIFENIVTNNAN 371
Db 296 LTLASGGGGISFSSNNIVQGTAGNGAIGSLAAGCSLSAEAGDITFNGNAL----- 348
OY 372 GTTSANPPRRNITAYASSGELLGAGSSON--LFDPIFVSAGVS---VSFNK----- 423
Db 349 -VATPQITKRNIDIGSTAKITNLRAISGHSIFEDPTIANTAAADSTDLNMLKADAGN 407
OY 424 EADOTGSVFSGATVNS-----ADFHORNLOTKTPAPLTSNGLFICEDHAOLTVNRFTO 478
Db 408 STDYSGSIYSGEKLSEDEKAVAD-----NLSTLKQPVYTLTACNLVLKRGVYLLDTGFTQ 463
OY 479 T-GGVVSLNGAVLSCYKNGAGSASNASITLKHIGLNTSLKSGAEIPLMVEPTNNS 537
Db 464 TAGSSVIMDAGTTL-----KASTEYTLGLSLPVDs-LGEKKKVI----- 504
OY 538 NNTYADTAATFSLDVKLS-----LIDDYGNSPRESIDL--THALSSQPLSISEASDNOL 591
Db 505 -----AASAASKNVALSGPILLLDNQGNA--YENDLKGTOQSFVSQSLGATATTDOV 556
OY 592 RSDMDPFGSLNVP-----HYGMOGLMTWGMMAKTODPEPASATITDPOKANFHTLL 645
Db 557 PA-----VPTVATPTHYVGOGTWGMKMWVDITSTPT-----KTATL 593
OY 646 TWLPAGVSPKHSPLANTLWGMN--LLATES--LKSNAELTPSDHPFWGITGGGLGM 702
Db 594 AMNTGYLPRPEROGPLVPSNLGWSFSDIOAIOGVIERSLTLCSDRGFV---AAGVANE 650
OY 703 VYDDPRENHGFHMRSSGYSAGMIACOT---HTFSLKFSQTYTKLWERYAKNNVSKNS 759
Db 651 LDKDKRGEKKRYKNGGVAIGG-AAOTGSENLISFAFOLFSGS---DKFELVAKNHT 704
OY 760 CGEEMFLSLOE-----GFLTTLVG-----LYSYGDHNCHEFT--QGEN 797
Db 705 DTYAGAFYIOHTIECSGFIGCLLDKLPGSMHKLPLVLEGLALSHVSNLDKTYTAYPEV 764
```

01-OCT-2000 (TREMblrel. 15, last annotation update)
DE POLYMORPHIC MEMBRANE PROTEIN A FAMILY.
GN PMP-19 OR CP0213.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CM1029;
RX MEDLINE=9920606; PubMed=10192288;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RT Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RL "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."; Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RA Shirai M., Hirakawa H., Ouchi K., Tabuchi M., Kishi F., Kimoto M.,
RA Takeuchi A., Nishida J., Shibata K., Fujinaga R., Yoneda H.,
RA Matsushima H., Tanaka C., Furukawa S., Miura K., Nakazawa A.,
RA Ishii K., Shiba T., Hattori M., Kohara S.;
RT "Comparison of outer membrane protein genes omp and pmp in the whole genome sequences of Chlamydia pneumoniae isolates from Japan and US."; Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brumham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Umayam L.A., Uterback T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA Bowman C., Dooson R., Gwin M., Nelson W., Deboy R., Kolonay J.,
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39."; Nucleic Acids Res. 28:1397-1406(2000).
RL EMBL: AE001638; AAD18679.1;
DR EMBL: AB033816; BA085967.1;
DR EMBL: AE002181; AAF38083.1;
DR HSSP: Q90121; IKPT.
DR TIGR: CP0213;
SQ SEQUENCE 947 AA; 103642 MW; 20CE1DEE1606DFF CRC64;

Query Match 11.5%; Score 586; DB 2; Length 947;
Best Local Similarity 25.2%; Pred. No. 1,1e-31;
Matches 250; Conservative 173; Mismatches 390; Indels 178; Gaps 47;
QY 14 LSGLAKEVPSRIFLMPNSVDPPTKESLSNKT---SLTGDTNHLNGLYMLNRYIIAIILOK 70
DB 30 LSGHSGEDLEFLTRSSSPRTYSLRMDIYCOFAGNSHKPGAAFLNLGDLFFINS 89
QY 71 TPNEGAAYITIDYLSFDFQKEGIYFAKNLTPESGAIYASPSPTVE-----IRDT 123
DB 90 TPL--AALTKFI--HLGARGAGLSESNT-----FKGLHVLLENMSMGVLT 137
QY 124 IGPVIFENNT--CCRPFTSSNPAAVNKIREGAIHAQN---LYINHHVDFGFMKNF 176
DB 138 SGDLSEFINNTSYLVCGNNISYGP-----GALLGCRKSKALFFRDNRGTIFLKNK 188
QY 177 SYVR-----GCAIATATFYVSENGSFLMDN--ICIQNTAGKGAIIAGT--SNF 226
DB 189 ANQOESHGCVGAASSISP-----GSPITFADMOELFQENEGELGAIINDGAIITF 242
QY 227 ESNNGDLFEINNACGAGAFSPICSLTGRNINIFYNNRCFKNVETASSEASDGAIV 286
DB 243 ENNFQTFSSFKASRGCAVSRICNLSOMKDTLF-----TKAAKVGCAIT- 291
QY 287 TTRLDVTGNRGIFTSNDITKNYGAIYAPVTVLVD--NGPTFYINIA--NNKGAIYIDG 344
DB 292 ADVHIRDCGKGIIVEEN--SATAGAIYAVNAVCIDINAGPVEFINNSALGLNGAIYMQA 350
QY 345 T-SNKSISADRAIIFNENIVTNV--TNANGTSTANPRRRAIYVASSGELLGAGSS 401

DB 351 TGSILRLHNOGDIEFCGNKVRKQFHSNINSTNFTN---NAITLQAPREFSISANCG 406
QY 402 QNIIFFDPI---EVSNAQSVSFNKEADQTSVYSGATVSADPHONLDTKP---A 454
DB 407 HRCIEFDPIISATENYNS--LYINHQRLEAGCAVIFSGARLSE--HKENKNKTSIINO 463
QY 455 PLTSLNGFLCIEDHQAQTVNRFOTGGVYSLNGAVLSCYKNGAGNSASNAITLKHGL 514
DB 464 PVLCSGLVSIIEGAILAVRSEFGGLLALPGSKULT---QGNSEKDKIVITNLGE 519
QY 515 NLSSILKSG-AEPLILMVEPTNNNNYTDATATPSLSYKL-----SLID--DYGNSP 565
DB 520 NLENLSDSPAET-----RATEKASIEISGVPRVYCHTESFYEHEHYSKRP 565
QY 566 YESTDLTHALSSQPMILISEASDNLRS---DDMFSGLVNPHYGMGLMTGMAKTDP 622
DB 566 YTTSI--LSAKKLVLTARSPEKQIQLIASEYMG-----YGOGSMEFSMSPNDTK 617
QY 623 EPASSATITDPOKANRFRHTLLTWLPAG--YVSPKRRSPLJANTLGMNL-----LA 674
DB 618 E-----KRTIISMTPTGEFSLDPRKRSFIPTLMTSFGSLNIANSIV 661
QY 675 TESLKMSALITSDHFRFNGITGGGLGMVYQDREHNPGRHSSGYSAGIACOTHTFS 734
DB 662 NNHYLNNESEVTPLOH--LCVFGPVYQIOMEONPKOSSNNLVOHAGHNVC--ARIPIFSFN 717
QY 735 LKFSQYTKLNERYAKNNYSKNYSCOGEMLF---SLOGFLTKLYGLYSYGDH---C 788
DB 718 TILSAALTLPLFSSSQNVADKSHA---QILGYSLNKSQALSRSSFSTEDSOYWK 774
QY 789 HHFYTGEMLTSGCTFRSOTMGCAVFFDLPMKPFQSTHILTAFLGALCIYSSL--SHT 846
DB 775 HVEPYKG---TSRGSWRNMGSGVSMYAY--PKGIRYLKMPFDLO--YTKLVQNPV 828
QY 847 EVCAYRSTSTKPLINLVPLGVKGSFNNATQRPDAW--TVELATQPVLYROEPIATOL 905
DB 829 ETGYDPRYFSS--SEMTNLSLPIGIALEMRFICRSLSFLQVSTSYIKDLRRVPOSSASL 887
QY 906 LASKGIFGSGSPSRHAMS-----YKI 928
DB 888 VLNHYTWMDIQGVPLCKEALNITLNSTIKYKI 918

RESULT 23
Q92398 PRELIMINARY: PRT: 928 AA.
AC Q92398;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE OUTER MEMBRANE PROTEIN PRECURSOR (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).
GN OMP10 OR PMP-9 OR CP0306.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CM1029;
RX MEDLINE=9920606; PubMed=10192288;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RT Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RL "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."; Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VR1310;
RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
RA Christiansen G., Birkegaard S.;
RT "Chlamydia proteins containing the GGA1-repeat belong to a subfamily of autotransporting pathogenicity factors."; Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.

QY 154 G-----AIIAONLXIIHNHNDVGFEMKNSFVSGAISTANTFVS--ENOSCFLEMDN 204
 DB 165 ASALOPTDLSIYEN--ISOSIKIFGFLNLF-----GSAISSSTAVAKFINNTATMSFSHN 218
 QY 205 ICIOTNTAGKGAIVAGTSNFSNCDLFFINNACC-----AGCAIFS 248
 DB 219 F-----TSSGGVIVGSSLLFENNNGCIIFTANCVSLSKGVTPSSGTALGSGAICI 273
 QY 249 P--ICSLTGNRCNIVFYNRRCFKNVETASSASDGAIKVTRLDVTCNRCRIFESDNT 306
 DB 274 PTGTFELKNNQCKCTFSYN-----GTPNDAGAIYAET-CNIVNOCALLDSDNTA 322
 QY 307 KNYGAIYAPVTVLDNCPTEYFINNANKGAIYI-----DGTNSKISADRHAI 358
 DB 323 ARNGAICAKVANIIOGRPIESRRRA-EKGAIFIGSVGDPKAKOTSTLTILASEGIA 381
 QY 359 FENNIIVTNTANGSTSTSNPRRAIVASSSGEILGAGSSONLIFEDPI-----E 411
 DB 382 FCGNML-----NTKPGIRNAIIVEAGGEIVLSAOGGSLVYEDPTTHSLPTTS 430
 QY 412 VSNAGVSFENKADQTSVVFSGATVNSAD-FHQRNLQTKTPAPLTLNSGFLCIEDHAQ 470
 DB 431 PSNKDITINAN---GASGVVFTSKGLSTELLIPANTTTLLGTVKIASGELKITDNAV 487
 QY 471 LTVNRF-TOTGVSLSGNGAVLSCYKNGAGNSASNAITLKHIGLNSLSKSGAEIFLL 529
 DB 488 VNVLFATOGSGQLTLGSGGTGLATPTGAPAA--VDFITGLADPSPFLKRD----- 539
 QY 530 WVEPTNNSNTYADTAATFSLSDVLSLIDYGNSPYESTDTHALSSQPMLSISEASON 589
 DB 540 FVSASVANG--TKNVTLTGAL-----VLDE-----HDVTDLYDNVSLQSPVAIPAIYAV-- 584
 QY 590 QLRSDMDPSGLNV-----PHYGQGLMTGMAK-TDPEPASSATITDPOK 635
 DB 585 -----FKGATVTKTGPDEIATPSHYGCKSYTMSHPLIPAPDGGFPGSPSPS 636
 QY 636 ANRFRITLLTV-----LPAGYVPSKHSPLIANTLWGMMLATESLKNSE-----EL 684
 DB 637 AN---TLVAWNSDTLVRSTYILDPERYGEIVNSILW-----ISFLGNOAFSDIIOV 686
 QY 685 TPDSHPFGITGGGLGMYVODPRENHPGFHMRSSYSAGMIAGQTH--TFSLSKSGQYTT 742
 DB 687 LLDHPGLSTAKALGAVENHPRGHGFSGRYGQYQALSMNTDHTTGLSTGSLYXG 746
 QY 743 KLNERYAKNNVSKNY--SCQEMFLSLOEGFLTLKVLGYLYCYGCHNHFFYQGENL-T 799
 DB 747 KTNAPYDSCSEQWYLLSPFGOPPIVTOKSEALISMKAACYSKNHLNTTYLRDPAK 806
 QY 800 SOGTFRSOTMGAVFEDLPKMPFGSTHILAP-----FLGALGIYSSLSHTEVGA 850
 DB 807 SOGOMHNNSYVYLISAE--HFELMWCILTRLAQAMDLSGFIASEFLGQWQSKTEFGD 863
 QY 851 YPRSFSTPLINLVPIYGVKSGFPMNATOR--POAWTVELAVOPVLYROEPIATOLLASK 909
 DB 864 LORGS-RGKGYNSLPLGCSQWFTPRKAPSTLTILKAYRPDIYRVAPHIVIVVSMQ 922
 QY 910 GIWFGSGSPSRHANSYKISOOTPLSLWTLFLFOYHGFYSSTPCNY 956
 DB 923 ESTSISGANLRHGLFVQI-HDVVDLTEDTQAFNLVTFEGKNGFTNH 968
 RESULT 25
 Q9RB63 PRELIMINARY: PRT: 978 AA.
 AC Q9RB63:
 DT 01-MAY-2000 (Tremblrel. 13. Created)
 DT 01-MAY-2000 (Tremblrel. 13. Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 13. Last annotation update)
 DE POLYMORPHIC OUTER MEMBRANE PROTEIN H FAMILY (POLYMORPHIC MEMBRANE
 DE PROTEIN H FAMILY).
 DE PMP 14 OR CP0298.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria: Chlamydiales: Chlamydiaceae: Chlamydia.

OX NCBI_TaxID=83558:
 RN (1)
 RC SEQUENCE FROM N.A.
 RC STRAIN-J138:
 RA Shiral M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.,
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CWL029 from USA."
 RL Nucleic Acids Res. 28:2311-2314(2000).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-AR39:
 RX MEDLINE=20150255; PubMed=10684935.
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Uterback T.,
 RA Berry K., Bass S., Linher K., Weidman J., Khouli H., Craven B.,
 RA Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J.,
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.,
 RT "Genome sequences of Chlamydia trachomatis Moyn and Chlamydia
 RT pneumoniae AR39."
 RL Nucleic Acids Res. 28:1397-1406(2000).
 DR EMBL: AP002546; BAA98661.1;
 DR EMBL: AE002191; AAF38155.1;
 DR TIGR: CP0298;
 DR INTERPRO: IPR000515;
 DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN_1;
 SO SEQUENCE 978 AA; 103654 MW; 6EE142999D3019 CRC64;
 Query Match 11.5%; Score 583; DB 2; Length 978;
 Best Local Similarity 26.1%; Pred. No. 1.8e-31;
 Matches 263; Conservative 137; Mismatches 411; Indels 196; Gaps 45;
 QY 41 SNKISLTGDTFNL--TNCYL-DNLRYIIAIILOKTPNEGAATYDYLSPFOKEGIYFA 97
 DB 67 SSNLSLKGKSLTFTSOQAPTNVALL-----SAAETLT-FKNF-----SSINFT 112
 QY 98 KNLTPESGCAIGYASPNSTPYEIRDTGPIVFEEN--TCCRPTSSNPNAANKIREG 153
 DB 113 GNSOTGLGCLL-YGKD-----IVFOSIKDLIFTRRVAVSPASVTSATP-AITVTG 164
 QY 154 G-----AIIAONLXIIHNHNDVGFEMKNSFVSGAISTANTFVS--ENOSCFLEMDN 204
 DB 165 ASALOPTDLSIYEN--ISOSIKIFGFLNLF-----GSAISSSTAVAKFINNTATMSFSHN 218
 QY 205 ICIOTNTAGKGAIVAGTSNFSNCDLFFINNACC-----AGCAIFS 248
 DB 219 F-----TSSGGVIVGSSLLFENNNGCIIFTANCVSLSKGVTPSSGTALGSGAICI 273
 QY 249 P--ICSLTGNRCNIVFYNRRCFKNVETASSASDGAIKVTRLDVTCNRCRIFESDNT 306
 DB 274 PTGTFELKNNQCKCTFSYN-----GTPNDAGAIYAET-CNIVNOCALLDSDNTA 322
 QY 307 KNYGAIYAPVTVLDNCPTEYFINNANKGAIYI-----DGTNSKISADRHAI 358
 DB 323 ARNGAICAKVANIIOGRPIESRRRA-EKGAIFIGSVGDPKAKOTSTLTILASEGIA 381
 QY 359 FENNIIVTNTANGSTSTSNPRRAIVASSSGEILGAGSSONLIFEDPI-----E 411
 DB 382 FCGNML-----NTKPGIRNAIIVEAGGEIVLSAOGGSLVYEDPTTHSLPTTS 430
 QY 412 VSNAGVSFENKADQTSVVFSGATVNSAD-FHQRNLQTKTPAPLTLNSGFLCIEDHAQ 470
 DB 431 PSNKDITINAN---GASGVVFTSKGLSTELLIPANTTTLLGTVKIASGELKITDNAV 487
 QY 471 LTVNRF-TOTGVSLSGNGAVLSCYKNGAGNSASNAITLKHIGLNSLSKSGAEIFLL 529
 DB 488 VNVLFATOGSGQLTLGSGGTGLATPTGAPAA--VDFITGLADPSPFLKRD----- 539
 QY 530 WVEPTNNSNTYADTAATFSLSDVLSLIDYGNSPYESTDTHALSSQPMLSISEASON 589
 DB 540 FVSASVANG--TKNVTLTGAL-----VLDE-----HDVTDLYDNVSLQSPVAIPAIYAV-- 584

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 18:41:45 ; Search time 34.43 Seconds
(Without alignments)
538,440 Million cell updates/sec

Title: US-09-677-752-2

Perfect score: 5086

Sequence: 1 MKKAFFFLIGNSLGLARE.....GFTSSSTFCNVLNGELALRF 965

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCRTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	194	3.8	1612	1	US-08-169-927-2
2	177.5	3.5	1338	2	US-08-728-470-9
3	177.5	3.5	1599	2	US-08-617-697-9
4	177	3.5	1477	1	US-08-038-682-4
5	177	3.5	1477	1	US-08-302-832-4
6	177	3.5	1477	2	US-08-530-198-4
7	177	3.5	1477	2	US-08-469-880-4
8	177	3.5	1477	2	US-08-728-470-4
9	177	3.5	1477	2	US-08-617-697-4
10	176.5	3.5	1536	1	US-08-038-682-2
11	176.5	3.5	1536	1	US-08-302-832-2
12	176.5	3.5	1536	2	US-08-530-198-2
13	176.5	3.5	1536	2	US-08-469-880-2
14	176.5	3.5	1536	2	US-08-728-470-2
15	176.5	3.5	1536	2	US-08-617-697-2
16	174	3.4	1160	4	US-08-808-599A-24
17	171.5	3.4	1529	2	US-08-728-470-10
18	171.5	3.4	1600	2	US-08-617-697-10
19	154.5	3.0	1026	1	US-08-194-290-7
20	150	2.9	983	2	US-08-164-292B-26
21	150	2.9	983	3	US-08-845-623-26
22	150	2.9	983	3	US-08-815-927-26
23	146.5	2.9	1026	2	US-08-614-377A-7
24	146.5	2.9	1026	4	US-09-142-648B-7
25	146	2.9	1007	4	US-08-551-459-4
26	144	2.8	1912	1	US-08-409-995-4
27	144	2.8	1912	3	US-08-685-467-4

28	144	2.8	2353	4	US-09-377-155-33	Sequence 33, Appl
29	144	2.8	2353	4	US-08-913-942-4	Sequence 4, Appl
30	143	2.8	890	2	US-08-483-101-14	Sequence 14, Appl
31	139	2.7	1394	5	PCT-US95-10661A-2	Sequence 2, Appl
32	138	2.7	1026	2	US-08-542-003-6	Sequence 6, Appl
33	138	2.7	1026	2	US-08-322-760A-6	Sequence 6, Appl
34	133.5	2.6	820	1	US-08-291-896-2	Sequence 2, Appl
35	133.5	2.6	820	2	US-08-485-278-2	Sequence 2, Appl
36	133	2.6	792	3	US-08-433-522A-8	Sequence 8, Appl
37	133	2.6	792	3	US-09-135-166-8	Sequence 8, Appl
38	131	2.6	704	3	US-08-792-832A-2	Sequence 2, Appl
39	131	2.6	797	3	US-08-433-522A-2	Sequence 2, Appl
40	131	2.6	797	3	US-08-433-522A-4	Sequence 2, Appl
41	131	2.6	797	3	US-08-433-522A-6	Sequence 6, Appl
42	131	2.6	797	3	US-09-135-166-2	Sequence 2, Appl
43	131	2.6	797	3	US-09-135-166-4	Sequence 4, Appl
44	131	2.6	797	3	US-09-135-166-6	Sequence 6, Appl
45	129	2.5	666	2	US-08-737-716-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-08-169-927-2
Sequence 2, Application US/08169927
Patent No. 5783441
GENERAL INFORMATION:
APPLICANT: Carl, Mitchell
APPLICANT: Dobson, Michael E.
APPLICANT: Ching, Wei Mei
TITLE OF INVENTION: Gene and Protein Applicable to the
TITLE OF INVENTION: Preparation of Vaccines for Rickettsia typhi and the Detection of Both
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Counsel, Naval Medical R & D Command
STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
CITY: Bethesda
STATE: MD
COUNTRY: USA
ZIP: 20889-5606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/169,927
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/742,128
FILING DATE: 08/09/91
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, A. David
REGISTRATION NUMBER: 24,743
REFERENCE/DOCKET NUMBER: 75,976
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-6759
TELEFAX: (301) 295-1022
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1612 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-169-927-2
Query Match 3.8%; Score 194; DB 1; Length 1612;
Best Local Similarity 18.9%; Pred. No. 2.3e-07;

```

OY 332 IANNKGA--IYIDGTSNKSISADRHAIIFENIYV-----NVTNANGT-----ST 375
Db 807 ASGNADAKKVTFFKVKXKSIKTDGHNVTLNSEVKTSGSSNAGNDNSTGLTISKADVTV 866
OY 376 SANPRRRAIVTAVSSGELLGAGSSONLIFYDPLEVSNAVSVSFNKEADOTGSVFSG 435
Db 867 NNNVTSKTTINISAAAGVNTTKEGT-----INATGSEVY-AQNGTIKGNITSON 917
OY 436 ATVSADFHQRMLOKTPAPPLTSLNGFLCIEDHAQITVNRFTQGTGVVSLGNGAVLSCYK 495
Db 918 VLVATE-----NLVTENAVINATSG-----TVNISKTGDI-----K 951
OY 496 NGAGNSASNASITLKHIGLNLSSILKSGAEIPL-----WVEPTNNSNNYTD 543
Db 952 GGIESTSGNVNITASGNLTKVSNL--TGODVTVTADAGALTTTACSTISATIGNNITTK 1009
OY 544 T-----AATFSLDVKLSLIDYGNL-----PYESTDLTHALSQPMPLSISEASD 588
Db 1010 TGDINGKVESSGSVTLVATGATLAVGNISGNTVTTADSGKLTSTVGSTINGTNSVTT 1069
OY 589 NOLSRDDMDFG--LNVPHYGMOGLMTWGMKTODPEPASSATITDPOKANRFRHTLL-- 644
Db 1070 SOSGIEGTISGNTVNTV-----ASTGDLTIGNSAKV-----EAKNGAATLTAE 1113
OY 645 ---LTLWPAQVVPSPKHSPLIA--NTLWGNMLATESLKNSAEITPSDHPWGTITGGL 699
Db 1114 SGKLTITQGSSTISSNGOTTLTAKDSIAGININAAVTLNTGTLTIGTDSKINATSGTL 1173

RESULT 3
US-08-617-697-9
: Sequence 9, Application US/08617697
: Patent No. 5977336
: GENERAL INFORMATION:
: APPLICANT: Barekamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Matlare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/617,697
: FILING DATE: 01-APR-1996
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 05-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Beikstieser, Jeffrey W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-557
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1599 amino acids
: TYPE: amino acid
: STRANDEDNESS: single

```

```

: TOPOLOGY: linear
US-08-617-697-9

Query Match 3.5%; Score 177.5; DB 2; Length 1599;
Best Local Similarity 21.4%; Pred. No. 6e-06;
Matches 154; Conservative 93; Mismatches 266; Indels 187; Gaps 33;

OY 75 GAATITDLYLSEFFDQKEGIYFAKNLTPESSGALGASPNSEFYEIRDTGPIVFENNTC 134
Db 806 GGNITITQVEGTDSRVNKGVAKKNIT-FKGNITFGSOKA-TTEIK--GNVTINKNTN 860
OY 135 CRFTSSNPNAAVNR-----IRCGAI-HAONLIT-----NHHMD 168
Db 861 A---TLRGANFAENKSPINAGNVINNGNLTTAGSIINAGNLTVSKGANLQAITNTYFN 917
OY 169 VVGFV-----KFEYVVRGA-----ISTANTFVSENGSCFLEMNICIOTNAGKGAIT 219
Db 918 VAGSEFNNCASNISIARGAKFKDINNTSSLNITNSDTTY---RTILKGNISKSG--- 971
OY 220 AGTSNSESNNCDLFFINNACCA---GGAIFSPICSLTGNKGNIVFYNNRCFK-NVE-- 272
Db 972 -----DLNIDKSDAEIQIGNISKEGNLTISDKVNITNQTITKAGVEGG 1019
OY 273 -TASSEASDCAIKVTRLDVTGNRGRIFPSDNTITKNYGAIYAPVTVLVNCPPTFTNN 331
Db 1020 RSDSSEAEANLTIOTKELKLAGD-----LNISGFKAET--AKNGSDLTIGN 1066
OY 332 IANNKGA--IYIDGTSNKSISADRHAIIFENIYV-----NVTNANGT-----ST 375
Db 1067 ASGNADAKKVTFFKVKXKSIKTDGHNVTLNSEVKTSGSSNAGNDNSTGLTISKADVTV 1126
OY 376 SANPRRRAIVTAVSSGELLGAGSSONLIFYDPLEVSNAVSVSFNKEADOTGSVFSG 435
Db 1127 NNNVTSKTTINISAAAGVNTTKEGT-----INATGSEVY-AQNGTIKGNITSON 1177
OY 436 ATVSADFHQRMLOKTPAPPLTSLNGFLCIEDHAQITVNRFTQGTGVVSLGNGAVLSCYK 495
Db 1178 VLVATE-----NLVTENAVINATSG-----TVNISKTGDI-----K 1211
OY 496 NGAGNSASNASITLKHIGLNLSSILKSGAEIPL-----WVEPTNNSNNYTD 543
Db 1212 GGIESTSGNVNITASGNLTKVSNL--TGODVTVTADAGALTTTACSTISATIGNNITTK 1269
OY 544 T-----AATFSLDVKLSLIDYGNL-----PYESTDLTHALSQPMPLSISEASD 588
Db 1270 TGDINGKVESSGSVTLVATGATLAVGNISGNTVTTADSGKLTSTVGSTINGTNSVTT 1329
OY 589 NOLSRDDMDFG--LNVPHYGMOGLMTWGMKTODPEPASSATITDPOKANRFRHTLL-- 644
Db 1330 SOSGIEGTISGNTVNTV-----ASTGDLTIGNSAKV-----EAKNGAATLTAE 1373
OY 645 ---LTLWPAQVVPSPKHSPLIA--NTLWGNMLATESLKNSAEITPSDHPWGTITGGL 699
Db 1374 SGKLTITQGSSTISSNGOTTLTAKDSIAGININAAVTLNTGTLTIGTDSKINATSGTL 1433

RESULT 4
US-08-038-682-4
: Sequence 4, Application US/08038682
: Patent No. 5549897
: GENERAL INFORMATION:
: APPLICANT: BAREKAMP, STEPHEN J
: APPLICANT: ST. GEME III, JOSEPH W
: TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
: TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Matlare, Ltd
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia

```

```

Db 648 TRKNTSYWQTHSDHWN/SALNLETGANFTFIKYSNSKGLTTOYRSSAGV----- 699
QY 135 CRPTSSNPNAVNKIRREGAIIHAO-----NLXI 163
Db 700 --NFGVNGNMSFN-LKEGAKVNFELKPNENNTSKPLPIRFLANITATGGSVFFDIYA 756
QY 164 NHN-----HDVGFPMKFSYVRGAI-----STANTFVVSQSCFL 200
Db 757 NHSGGAEIKMSEINISGANCFTLN-SHYRGDAFKINKDLTINATNSNFSLRQTKDDFY 815
QY 201 --FMDNICIQT-NTAGKGAIIYAGTSNFSFNCDLFF-----INNA----- 239
Db 816 DGYARNAINSTYINISILGCVNTLGGONSSSITGNITTEKAAVNTLEANNAPNOONIRDR 875
QY 240 -CCAGGAIFSPICSLTGN-----RCNIVFYNNRCFK-----NVTASSEASD 280
Db 876 VIKGLSLVNGSLSLTGENADIKGNLTISESATFKGKTBDTINIGFTNGTAEINITO 935
QY 281 G-----GAIKVTTRL-----DVTGNRGRIFFSD----- 303
Db 936 GVVKLGNTNDDGLNITTHAKRNORSIIIGDIIINKGSLNITDSNDAEIQIGNISQKE 995
QY 304 -----NITK-----NYGGAIIYAPVTLVD----- 322
Db 996 GNLTISSDKINITKOITIKKIDGEDSSDATSNANLTIKTELKLTEDLSISGFNKAEI 1055
QY 323 ---NGPTYFINNIANNKGA---IYIDGTSNKSISADRHAIIFENIYVNTYNNANGTST 375
Db 1056 TAKDRODLTIGNSNGNSGAELKAYTFNNVKDSKISADCHNTLNSKYVTSSNGSGREN 1115
QY 376 SANPRRNAITYA-----SSGGEILLGAGSSQNLIFYDPIEVSNAGV 417
Db 1116 SOND---TGLITTAKNVEYNKDITSLKTIVNITASEKVTTTAAST-----INATNGKA 1164
QY 418 SVSFNKADQTSVVSFGATVN-SADFHORNLOTKTPAPLLISNGFLCIEDHAOL----- 471
Db 1165 SIT-TKTGDISGTI--SGNTVSASATV--DLTTSKSGITEKSGEANTVSATGTIGTI 1218
QY 472 ---TVNRFOTOGVYSLGCAVILSCYKNGAGNSASNASITLKHIGLNLSSILKSGAEIPL 528
Db 1219 SCNTYV-NVANGDLTVGCAEINATEGATLTLATGNLTITE-AG--SISITKQOVDL 1273
QY 529 L---WVEPTNNSNNTADTAATFSL---SDVKL---SLIDD-----YGNPFYESTDLT 572
Db 1274 LAONGSIAGSIANAANTLMTGTLTIVAGSDIKATSGTLVINAKDAKMLNGDASCDSTEY- 1332
QY 573 HALLSOPMLISEASON--QLRSDMDPESGLNV 603
Db 1333 NAWNAGSGSVTAATSSVNTIGDLNTVNGLNI 1365

```

RESULT 6
US-08-530-198-4
Sequence 4, Application US/08530198
Patent No. 5869065

GENERAL INFORMATION:
APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GEME III, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Maltare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentlin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530,198
FILING DATE: 13-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERNSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: JWB-1186
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-530-198-4

```

Query Match 3.5%; Score 177; DB 2; Length 1477;
Best Local Similarity 19.4%; Pred. No. 5.8e-06;
Matches 158; Conservative 115; Mismatches 244; Indels 296; Gaps 40;

```

QY 37 KESLSNKSISLTD-----THNLTCYLDNLRILALIQKTPNEGAAVTITD- 82
Db 603 KOFRRANNVSLNKGTLGKGLNISSVNNLTHLSG-----TINISGNTIINTQ 647
QY 83 ---YLSFDTQEGEYFAKNLTPESGG---AIGVSPNSP--TVEIRDTGPIVENNTC 134
Db 648 TRKNTSYWQTHSDHWN/SALNLETGANFTFIKYSNSKGLTTOYRSSAGV----- 699
QY 135 CRPTSSNPNAVNKIRREGAIIHAO-----NLXI 163
Db 700 --NFGVNGNMSFN-LKEGAKVNFELKPNENNTSKPLPIRFLANITATGGSVFFDIYA 756
QY 164 NHN-----HDVGFPMKFSYVRGAI-----STANTFVVSQSCFL 200
Db 757 NHSGGAEIKMSEINISGANCFTLN-SHYRGDAFKINKDLTINATNSNFSLRQTKDDFY 815
QY 201 --FMDNICIQT-NTAGKGAIIYAGTSNFSFNCDLFF-----INNA----- 239
Db 816 DGYARNAINSTYINISILGCVNTLGGONSSSITGNITTEKAAVNTLEANNAPNOONIRDR 875
QY 240 -CCAGGAIFSPICSLTGN-----RCNIVFYNNRCFK-----NVTASSEASD 280
Db 876 VIKGLSLVNGSLSLTGENADIKGNLTISESATFKGKTBDTINIGFTNGTAEINITO 935
QY 281 G-----GAIKVTTRL-----DVTGNRGRIFFSD----- 303
Db 936 GVVKLGNTNDDGLNITTHAKRNORSIIIGDIIINKGSLNITDSNDAEIQIGNISQKE 995
QY 304 -----NITK-----NYGGAIIYAPVTLVD----- 322
Db 996 GNLTISSDKINITKOITIKKIDGEDSSDATSNANLTIKTELKLTEDLSISGFNKAEI 1055
QY 323 ---NGPTYFINNIANNKGA---IYIDGTSNKSISADRHAIIFENIYVNTYNNANGTST 375
Db 1056 TAKDRODLTIGNSNGNSGAELKAYTFNNVKDSKISADCHNTLNSKYVTSSNGSGREN 1115
QY 376 SANPRRNAITYA-----SSGGEILLGAGSSQNLIFYDPIEVSNAGV 417
Db 1116 SOND---TGLITTAKNVEYNKDITSLKTIVNITASEKVTTTAAST-----INATNGKA 1164
QY 418 SVSFNKADQTSVVSFGATVN-SADFHORNLOTKTPAPLLISNGFLCIEDHAOL----- 471
Db 1165 SIT-TKTGDISGTI--SGNTVSASATV--DLTTSKSGITEKSGEANTVSATGTIGTI 1218
QY 472 ---TVNRFOTOGVYSLGCAVILSCYKNGAGNSASNASITLKHIGLNLSSILKSGAEIPL 528

```

APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstesser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-4

Query Match 3.5%: Score 177; DB 2; Length 1477;

Best Local Similarity 19.4%; Pred. No. 5.8e-06;
Matches 158: Conservative 115; Mismatches 244; Indels 296; Gaps 40;

QY 37 KESLSNKLSTLGD-----THNLNCLYLDNLRYLAILAOKTPNEGAATVITD- 82
DB 603 KDFRANNVSLNCTGKGLNISSVNNLTHNLSC-----TINISGNTITNOT 647
QY 83 ---YLSFDTQKEGIFYFAKNLTPESGC---AIGVAPNSP--TVEIRDTGPVIFENNTC 134
DB 648 TRKMTSYWOTSHDSHMVNSALNLETGANFTIKYISSNSKGLTQYRSSAGV----- 699
QY 135 CRPTSSPNAVAKIRREGCAIHAQ-----NLXI 163
DB 700 --NENGVAHNSFN-LKGAIVNFKLPKNENMNTSKPLRIFLANITATGGGVFEETIYA 756
QY 164 NNN-----HNVGFMKNFSYVRCGAI-----STANTFVSENOGFL 200
DB 757 NMSGRAELKSEINISNCAFTLN-SHVRGDDAFKINKDLTINATNSFSLRQKDFY 815
QY 201 --FMDNICIQT-NTAGKGAIVAGTSNFSFESNCDLFF-----INNA----- 239
DB 816 DGYARNAINSTYINISILGCVTLTGONSSSITGNTITEKAAVTLNANNAPOQNIHDR 875
QY 240 -CCAGAGAFSPICSLTGN-----RGNIFYNNRCFK-----NVEETASEASD 280
DB 876 VIKUGSLLVNGLSLTGENADIKGLTISEATFEGKTRDRLNLTGNTNGTAELINITQ 935
QY 281 G-----GAIVTTRL-----DVTGNRGRIFFSD----- 303
DB 936 GVVKLGANTNGDNLITTHAKRNRKSIIGDILINKKGSILNDSNDAEIQIGNISQKE 995
QY 304 -----NITK-----NYGAIYAPVTLVD----- 322
DB 996 GNLITISDKINITKOITIKKIGDGEDSSSDATSNANLTIKTEKLTEDLSISGFNKAEI 1055
QY 333 ---NCPYFINNIANNKGA---TYIDGTSNKSISADRHAIIFENMIVTNTANGST 375
DB 1056 TAKDREDLTLIGNSDNGSGAEAKIVTFPNVVKDSKISADGHVTLNLSKRYKTSNGGREGSN 1115
QY 376 SANPRRAATVA-----SSGCEILLGAGSSONLIFYDPIEVSAGV 417
DB 1116 SDND---TGLTITAKNVEVKNQDITSLKTYNITASEKVTJTAST-----INATNCGA 1164
QY 418 SVSFNKADQIGSVFSGATVNA-SADFHQRNLQTKTPAPRLTSLNGFLCIEDHAOL----- 471

DB 1165 SIT-TKGTDISGTL--SGNTVYSATV---DLTKSGSKIEAKGEMVNSAGTIGCTI 1218
QY 472 ---TVNRFYQGVVSLGNGAVLSCKYRNGAGNSASNAITLKHIGLNSILKSGAIEPL 528
DB 1219 SGNTVN-VTANAGDLTYGNGAEINATEGAATLTATGNTLTTE-AG---SITSTKGGVDL 1273
QY 529 L---WVEPTNSNNYADTAATFSL---SDVKL---SLDD-----YGNSPRESTDLT 572
DB 1274 LAONGSIAGSINAAVTLNTGTLTTVAAGSDIKATSGTLVINAKDARLNGASGDSDEV- 1332
QY 573 HALSSQPLMSISEASDN--QLRSDMDSEGLNV 603
DB 1333 NAAVNASGSGSVTAATSSVNITGDLNVTYNGLNI 1365

RESULT 9
US-08-617-697-4

Sequence 4, Application US/08617697
Patent No. 5977336
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: Of No. 5977336-Typeable Haemophilus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstesser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-617-697-4

Query Match 3.5%: Score 177; DB 2; Length 1477;

Best Local Similarity 19.4%; Pred. No. 5.8e-06;
Matches 158: Conservative 115; Mismatches 244; Indels 296; Gaps 40;

QY 37 KESLSNKLSTLGD-----THNLNCLYLDNLRYLAILAOKTPNEGAATVITD- 82
DB 603 KDFRANNVSLNCTGKGLNISSVNNLTHNLSC-----TINISGNTITNOT 647
QY 83 ---YLSFDTQKEGIFYFAKNLTPESGC---AIGVAPNSP--TVEIRDTGPVIFENNTC 134

Sequence 10, Application US/08728470
Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0810
INFORMATION FOR SEQ. ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-10

Query Match 3.4%; Score 171.5; DB 2; Length 1529;
Best Local Similarity 23.5%; Pred. No. 1.8e-05;
Matches 104; Conservative 70; Mismatches 186; Indels 83; Gaps 21;
DB 139 TSSNPNAAVKKIREGAIHAQNLINHN---HDVYGFKNFVSYRGGAISTANTFVSEN 195
1033 TSNSSNAGNDNSTGLTISAKDVTNNNTSHKTI---MISAAGNVTTKEGTTINATT 1088
196 OSCFLFMDNICIOTNTAGKGAIIYAGTSNFSFESNCDLFIIN-----NCCAGGAIFSP 249
1089 GSVEVTAQNGTIGNITSONVTVTA--TENLVTTENA---VINATSGTVNISTKTGDIKGC 1144
250 ICSLTGN-----RGNIVFYNNRCFKNVETASSSEASDGAIKVTRLDVTGNRGRIFESDN 304
1145 IESTSGNVNTASGNTLKVSNIITGQV---TYADAGALTTTAGTISATG-----NAN 1196
305 ITKNYCG---GAIYAP--VTLVDNGPTFTYINNINANNKGAIIYDGTISNSKISADRAHAIIF 359
1197 ITTKTGIDINKVESSSGSLVLTGATLAVGNISGN-----TVTITADSGKLTS 1245
360 NENIVTNTANGTSTANSANPR-----RNATIVASSSGEILLGAGSSONLFIYDIETVSN- 414
1246 TVGSTINGTNSVTTSOSGDIETISGNIVNVTASTGDTLTIGNSAK-----VEAKNG 1297

DB 415 -AGVSVSFNKEADOTGSVFS--GATVNSA-----DPHQRNIQTKTPAPLTLNSGFLC 464
1298 AATLTAESECKLTTOTGSSITSSNGOITLLTAKDSSIAGNINAAVNTLUTLTGTTTDSKI 1357
465 IEDHAQITV-RTOTGCGVAVSLNGAVLSCYKNGAGN---SASMSIT--LKHI-GLNL 516
1358 NATSGTLTINAKAKLDGASGDRYVATNATNAGSGNVTAKTSSSVNITGDLNLTINGLNI 1417
517 -----SILKSGAIEIPLWVEP 533
1418 ISENGRNTVRLRGEKIDVYIOP 1440

RESULT 18
US-08-617-697-10
Sequence 10, Application US/08617697
Patent No. 5977336
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0810
INFORMATION FOR SEQ. ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1600 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-617-697-10
Query Match 3.4%; Score 171.5; DB 2; Length 1600;
Best Local Similarity 23.5%; Pred. No. 2e-05;
Matches 104; Conservative 70; Mismatches 186; Indels 83; Gaps 21;
DB 139 TSSNPNAAVKKIREGAIHAQNLINHN---HDVYGFKNFVSYRGGAISTANTFVSEN 195
1103 TSNSSNAGNDNSTGLTISAKDVTNNNTSHKTI---MISAAGNVTTKEGTTINATT 1158
196 OSCFLFMDNICIOTNTAGKGAIIYAGTSNFSFESNCDLFIIN-----NCCAGGAIFSP 249
1159 GSVEVTAQNGTIGNITSONVTVTA--TENLVTTENA---VINATSGTVNISTKTGDIKGC 1214

Mon May 7 09:06:32 2001

us-09-677-752-2.rai

Page 17

CITY: Boston
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/614,377A
 FILING DATE: 12-MAR-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/194,290
 FILING DATE: 09-FEB-1994
 CLASSIFICATION: 435
 APPLICATION NUMBER: US 07/895,367
 FILING DATE: 09-JUNE-1992
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Tsao, Y. Rocky
 REGISTRATION NUMBER: 34053
 REFERENCE/DOCKET NUMBER: 08106/002001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-542-5070
 TELEFAX: 617-542-8506
 INFORMATION FOR SEO ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1026 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: Linear
 MOLECULE TYPE: protein

US-08-614-377A-7

Query Match	2.9%;	Score 146.5;	DB 2;	Length 1026;
Best Local Similarity	19.6%;	Pred. No. 0.0013;		
Matches 107;	Conservative 73;	Mismatches 245;	Indels 121;	Gaps 20

```

0Y      35  PTKESLSKRISLTDGTDHNLTCYLDNL-----RYLIALLOKPNKGAVITTDVLSFSDIQ  90
Dp      73  PSAAGLOFLVDSTNTNTDNLDAIYSKFAQENRFINFSLNLTAGAGATA-----PAAA  126
OY      91  KEGIFYAKNLTPESGGAIGTASPMSPYIEIDTG-----PVIFENNTCCRPFT  139
Dp     127  YTCVSAAOTVATAYDKIIGNAVATACAGDVAAAAPLISROANIDYLTAFVRAANT--PFT  183
OY     140  SSNPAAVNKIREGGAIHAQULYINHNHDVYGFMKNFYSYVG--GAIISANTFVVSQNSC  198
Dp     184  A-----AADIDLAVKRALIGTILNNAITYSGIGGYATATAMINDLSDG  228
OY     199  FLFMDNIC-----IOTNTAGKGAIFYACTSNSFESNNCDLFFINNACGAGAIIPSPI  250
Dp     227  ALSTDMAGVNLFTAYPSSVSGSLSLSTGTDLTGTANDTFVAGHV--AGAAITLYG  284
OY     251  CSLTGNRGNIFY-----NNRCFKNVEPTASSEASDGGAIKVTTLDTDYGNNGRI  299
Dp     285  DTLSCGACTDVLNMQAAAYTALPTGVTTISCIETMN--VTSGAATLTMTSSGVTGLTA-  340
OY     300  FFSNDITKNYCGAIFYAPVJT-----LVNDGPNFYFINNIAKKGAIYIDST-----  349
Dp     341  -----LNNITSGA--AQYVTAGAGONLTATTAAQAAANNVADGARNYVASTGYTSGTT  399
OY     346  --SNKSISADRRHAIIFNENIVTN--VTNANCTSTSANPPRNAITVYASSGCEILLGAGSS  401
Dp     394  VCANSAAGCTVSVAANSSTTTTGCAIAYTGCTAATVTAQTNAGNAVNTTLTQADYVITGNS  455
OY     402  ONLITYPDIEVSNAGVSS--FNKEADDTGVSVSGAIVNSADDEHQRNLQKTPAPLTL-  455
Dp     454  TTAVITVOTATAATACATVAGRVNCAVTTITDSAAASATTAGCI-----ATYTLG  501

```

OY 459 SNGEFCIEDHQLVNNRFTQTGGVVSLCNGCAVLSCYKMGAGNSASMSISITLKHIGLUSS 518
| | : | | : | | : | | :
Db 502 SEGATIIDSSALTFYN-LSGTGTSLGIGRAL-----TAPPTANLTILWVG 547

OY 519 ILKSQA 524
: : ||
Db 548 LTTTGA 553

RESULT 24
US-09-142-

```

: Sequence 7, Application US/09142648B
: Patent No. 6210948
: GENERAL INFORMATION:
: APPLICANT: Smit, John
: APPLICANT: Bingle, Wade H.
: APPLICANT: No. 6210948Bellini, John F.
: TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAT
: FILE REFERENCE: 08106/002002
: CURRENT APPLICATION NUMBER: US/09/142,648B
: CURRENT FILING DATE: 1999-03-30
: PRIOR APPLICATION NUMBER: PCT/CA97/00167
: PRIOR FILING DATE: 1997-03-10
: PRIOR APPLICATION NUMBER: US 07/614,377
: PRIOR FILING DATE: 1996-03-12
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 7
: LENGTH: 1026
: TYPE: PRT
: ORGANISM: Caulobacter crescentus
: US-09-142-648B-7

```

Query Match	2.9%	Score 146.5;	DB 4;	Length 1026;
Best Local Similarity	19.6%;	Pred. No. 0.0013;		
Matches 107;	Conservative 73;	Mismatches 245;	Indels 121;	Gaps 20;

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 19:09:45 ; Search time 63.9 Seconds
(without alignments)
1037.835 Million cell updates/sec

Title: US-09-677-752-2

Perfect score: 5086
Sequence: 1 MKKAFFFFLIGNSLGLARE.....GFYSSSTFCNLTNGELALRF 965

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_67:*\n1: p1r1:*\n2: p1r2:*\n3: p1r3:*\n4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4989.5	98.1	964	2	E71460
2	3648.5	71.7	976	2	F81722
3	1120.5	22.0	928	2	H72074
4	1060.5	20.9	952	2	D81593
5	1052	20.7	934	2	A72075
6	951.5	18.7	946	2	D81594
7	948.5	18.6	946	2	C72075
8	866.5	17.0	1034	2	F71460
9	793	15.6	1025	2	G81722
10	655.5	12.9	1013	2	G71460
11	645	12.7	936	2	C72078
12	644	12.7	936	2	B81591
13	620.5	12.2	987	2	H81722
14	620	12.2	930	2	A81591
15	617	12.1	930	2	D72078
16	597	11.7	928	2	D72077
17	597	11.7	949	2	F81591
18	595	11.7	1016	2	H71460
19	586.5	11.5	928	2	G81591
20	586	11.5	947	2	D72067
21	585.5	11.5	928	2	B72077
22	583	11.5	978	2	G72076
23	583	11.5	978	2	B81593
24	575	11.3	973	2	F72076
25	575	11.3	995	2	C81593
26	570.5	11.2	922	2	B72131
27	569.5	11.2	922	2	F81539
28	567	11.1	1276	2	C81591
29	564	11.1	983	2	A81723

30	562	11.0	1407	2	B72078	polymorphic outer
31	546.5	10.7	986	2	B81675	polymorphic membra
32	525.5	10.3	841	2	E71518	polymorphic membra
33	519.5	10.2	975	2	F71518	hypothetical prote
34	517	10.2	1723	2	E72067	polymorphic membra
35	517	10.2	1732	2	C81601	polymorphic membra
36	512	10.1	392	2	B72075	polymorphic outer
37	474.5	9.3	1609	2	H72013	polymorphic membra
38	470.5	9.3	878	2	B71460	probable outer mem
39	457	9.0	867	2	F81721	polymorphic membra
40	395	7.8	1751	2	G71518	hypothetical prote
41	389.5	7.7	1672	2	C81675	polymorphic membra
42	386	7.6	1531	2	H71468	probable outer mem
43	377.5	7.4	1460	2	D81675	polymorphic membra
44	358.5	7.0	1520	2	A81731	polymorphic membra
45	342	6.7	1770	2	A71517	hypothetical prote

ALIGNMENTS

RESULT 1
E71460
probable outer membrane protein E - Chlamydia trachomatis (serotype D, strain UM3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: E71460
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A:Reference number: A71570; MIMID:99000809
A:Accession: E71460
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-964 <ARN>
A:Cross-references: GB:AE001360; GB:AE001273; NID:93329342; PIDN:AA068467.1; PID:9332
A:Experimental source: serotype D, strain UM-3/Cx
C:Genetics:
A:Gene: pmpe

Query Match 98.1% Score 4989.5; DB 2; Length 964;
Best Local Similarity 98.3% Pred. No. 9.2e-299;
Matches 949; Conservative 7; Mismatches 8; Indels 1; Gaps 1;

QY	1	MKKAFFFFLIGNSLGLAREVPSRIPLMPNSVDPPTKESLSNKISLTGDTNLTNCYLDN	60
DB	1	MKKAFFFFLIGNSLGLAREVPSRIPLMPNSVDPPTKESLSNKISLTGDTNLTNCYLDN	60
QY	61	LRYLAILOKTPNCAAVTTIDYLSFFDQKEGYIFAKNLTPEGGAIGYASPSPTVEI	120
DB	61	LRYLAILOKTPNCAAVTTIDYLSFFDQKEGYIFAKNLTPEGGAIGYASPSPTVEI	120
QY	121	RDITGVIEFENNCCREPTSSNPAAVKKIREGAIHONLYIHNDVYGFMKNFYSVR	180
DB	121	RDITGVIEFENNCCREPTSSNPAAVKKIREGAIHONLYIHNDVYGFMKNFYSVR	180
QY	181	GCAISTANTFVVSNSQCFLEFMDNICIOTNTAGKGAIVYACTSNSPESNCDLFFINNAC	240
DB	181	GCAISTANTFVVSNSQCFLEFMDNICIOTNTAGKGAIVYACTSNSPESNCDLFFINNAC	240
QY	241	CAGGAIFSPICSLTGNRGNIVFYNNRCKNKTETASSESDGCAIKVTRLDVNTNRGRIF	300
DB	241	CAGGAIFSPICSLTGNRGNIVFYNNRCKNKTETASSESDGCAIKVTRLDVNTNRGRIF	300
QY	301	FSDNTKNGAIIAPVYTVLDNGPTTFINNANNGKAIIYIDGTSNKSISADRHAIIFN	360
DB	301	FSDNTKNGAIIAPVYTVLDNGPTTFINNANNGKAIIYIDGTSNKSISADRHAIIFN	360
QY	361	ENIVTNTNANGTSTANPPRRNATYVASSGGEILLGAGSSQNLIFYDPIEVSNAGSVS	420
DB	361	ENIVTNTNANGTSTANPPRRNATYVASSGGEILLGAGSSQNLIFYDPIEVSNAGSVS	420
QY	421	ENIVTNTNANGTSTANPPRRNATYVASSGGEILLGAGSSQNLIFYDPIEVSNAGSVS	480
DB	421	ENIVTNTNANGTSTANPPRRNATYVASSGGEILLGAGSSQNLIFYDPIEVSNAGSVS	480

A: Molecule type: DNA
 A: Residues: 1-938 <ARN>
 A: Cross-references: GB:AE001631; GB:AE001363; NID:94376750; PID:AA018608.1; PID:9437675
 A: Experimental source: strain CM1029
 R: Read, T.D.: Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gilm, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A: Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
 A: Reference number: AB1500; M0ID:20150255
 A: Accession: E81593
 A: Molecule type: DNA
 A: Residues: 1-938 <REA>
 A: Cross-references: GB:AE002190; GB:AE002161; NID:97189209; PID:AAF38143.1; PID:9718921
 A: Experimental source: strain AR39, HL cells
 C: Genetics:
 A: Gene: pmp_15; CP0286

Query Match 22.0%; Score 1120.5; DB 2; Length 938;
 Best Local Similarity 30.9%; Pred. No. 5e-61;
 Matches 308; Conservative 163; Mismatches 418; Indels 109; Gaps 31;

6 FFFLLGSLSLGAREVSRIFLPMNSVPDPTKESLSKISLTGDPHTLT-----VGFTHNQNDLAIVGNHD 59
 12 FTVVLAN--EGLQLPLETYITLSPEVOAQQ-----NC 56
 57 YLDNRYLALLOKTPNEGAATITDYLSFEDTOKEGYFAKNLTPESGAIGYASPSNP 116
 60 FILDYKY-----RSNGALGCKNLL--ISENIGVFEKKNVCPNSGAI-YAAQNC- 108
 117 TVEIRDTIGVIEENNTCCRPFTSSNNAVANKIREGAIHAONLNNHNDVYFKNF 176
 109 -----TISK--NQNYAFTNLVSDNPTATAGSL--GALFAIKCSTNNMGCGTFVNDL 159
 177 SYVKGAIATANFVSENSCFLFMDNICIQTNTAGKGAIGYAGTNSFESNCDLFEI 236
 160 ALNKGALYETNLSIDNKPILIKONRAL--NSDSIGGCIYSGNSLNEGNGAIOIT 217

237 NMACCAGGAIFS-PICSLTGNRCNIVEYNNCFKNVETASSEASDGAIAKVTRLDYTG 295
 218 SNSGSGGIFSTQTLTSSNKKLIEISENSAFAN--NYGSNFPGGGLTTTCTILNN 275
 296 RGRIFESDNTKNGGAIYAPVTLVNDGPTFYNNIANNKGAIYD-GTSNKK--ISA 352
 276 REGVLFNNOSQNGAIHAKSIILKEGYPVLTNATRGALLNSAGSGNSFLISA 335
 353 DRHAIFENENVTNANGSTSANPPRRAITVASSSGEILLAGAGSQNLFIYDPIEV 412
 336 DNGDIFNNN-----TASKHALNPYRNAIH-STPNMMLQIGARGYVLETDPIEH 386
 413 S-NGVSVSEKKEADOTGVSVFSGATVNSADFHORNIOTKTPAPLTLNGLIEDHAOL 471
 387 ELSPSFLIFEFEGHGTVLFSEGHVHONFTDEMNFSYLRNSELNGVLAVEDGAGL 446
 472 TVNRFOTGGVSLGNCAYVLCYKNGACN-----SASNAITLKATIGLNLKSG 523
 447 ACYFEGFGGTLILGOGAVITTT---AGTIPSPSTPTVTSITLNNIAIDLSLSEFQ 502
 524 AEIILWEPITNNNSNTADPAATFSLSDVLSLIDVGSNPSTEDTTHALSSOPMLSI 583
 503 AOAIRKIMYPTKGTSTIEDSNPTITISGT-LTLRNSNNEPDYBSLDSHSEKVPPLIYI 561
 584 SEASNDLRSDMDPFSGLNV-PHYGMOGLWTVGNAKTODPEPASSATITDPQK--ANRF 639
 562 VDVAQAOKINSQDLSTLNSGEHYGYOGIMSTYVET-----TTITNPTSLGANTK 613
 640 HRTLLTLVLPAGIYVSPKRRSPLANTLGNMLATBESLKNASALTLPDHPWGTG-- 697
 614 HKLLYANNSPLGYRHPRREREFITNALMQSAYTALAGLSLSS-----WDEEKSHA 665
 698 -----GLGMWVODPRENHPGFMHRSRGYSAGMIA--GOTHFSLKFSOTYTLNERYAKN 751
 666 ASLOGIGLLVHOKKNGCFGRSHMTGYSATTEATISSQSPNLSGFAQFSAKHESON 725

752 NVSSKNV-----SCGEMLFSLQEGFLLTKLVGL-XSYGDHNCCHFFYTOGENLTSOGERSO 807
 726 STSHHYFSGMCIENTLE--KENIRLS--VSLAYMTSEHTHTMYOGLGNSQGSFHHN 781
 808 TMGCAVFPDPMKPCFSTHILTPAFGLAGIYSSLSHTEVGAYPRPSFKTPILNVLVP 867
 782 TLGALSCVFLPQPHES-LQIYFPITALAIRGNLAFOGSDHAFESILHRLTIVSLP 840
 868 IGKVGSPMAATQRPQAMVELAYOPVLYRQEPQIATOLLASKGIMFGSGSPSSRHAMSK 927
 841 VGIASMKNNHRYPLVWLTEISYRSTLYRODPRLSHSLTISQGTWTQATPYVALGK 900
 928 ISOOTPLSLTLHPFOYHNGYSSSTCNLYNGELALRF 965
 901 VKNTMOVFPKVTLSLDYSADISSSTLSHLYNVASRMKF 938

RESULT 4
 DB1593
 polymorphic membrane protein E/F family CP0285 (imported) - Chlamydia pneumoniae
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #ext_change 11-May-2000
 C:Accession: DB1593
 R: Read, T.D.: Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gilm, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A: Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
 A: Reference number: AB1500; M0ID:20150255
 A: Accession: DB1593
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-952 <REA>
 A: Cross-references: GB:AE002190; GB:AE002161; NID:97189209; PID:AAF38142.1; PID:9718
 A: Experimental source: strain AR39, HL cells
 C: Genetics:
 A: Gene: CP0285

Query Match 20.9%; Score 1060.5; DB 2; Length 952;
 Best Local Similarity 30.6%; Pred. No. 2.5e-57;
 Matches 303; Conservative 164; Mismatches 447; Indels 77; Gaps 28;

6 FFFLLGSLSLGAREVSRIFLPMNSVPD-----PTKESLSK-----ISL 46
 8 FLYLGNFTACMGMTPAVYSLQDSLEKFALEDEEFRTSFLDLSLTLTGFSPTTF 67
 47 TGDTHNLT-NCYLDNRYLALLOKTPNEGAATITDYLSFEDTOKEGYFAKNLTPESG 105
 68 VGNRHNSSQDLVLSNYSIDNILLMTSAGAVACNNFL--LSNVEDHAFSSKALAGTG 125
 106 GAIGYASPSPTVEIRDTIGVIEENNTCCRPFTSSNNAVANKIREGAIHAQ-NLYIN 164
 126 GAIACQG---ACTITTKRGPILFFSNRGLN-----NASTGETRGCAIACNGDFTS 174
 165 HNHVYGFMKFSTYRGCAIANTAFVSENSCFLFMDNICIQTNTAGKGAIGYAGTNS 224
 175 ONOGTFEYVNNSVNMGALSTNGCHRLOSRAPLFEFN-----NTAPSGGALRENT 228
 225 SFESNCDLFEINNACGGAIFSPI-CSLTGNRCNIVEYNNCFKNVETASSEASDGA 283
 229 TISDNTRIYIKKNCNGNGGAIQTSYTAIKNNGSVYFENNNTALSG--SINSNGSGGA 286
 284 IKVTRLDYVNGRGRIFESDNTKNGGAIYAPVTLVNDGPTFYNNIANNKGAIYID 343
 287 I-YTNLISIDNPGTILFNNNYCIRHDGAICTOFILKNSGHVFTNN-QGMGAGMLL 344
 344 GTSNKSISADHAIIFNNIYNTNANGSTSANPPRRAITVASSSGEILLAGAGSQNN 403
 345 QDSCTCLLFAEGONIAFOKNEVFLTFG-----RYNAIHCTPNS-NIQLANNGYT 393
 404 LIFDPIEVSNAGVS-VSEKKEADOTGVSVFSGATVNSADFHORNIOTKTPAPLTLNGLF 462

A:Gene: CP0283

Query Match 18.7% Score 951.5; DB 2: Length 946;
Best Local Similarity 29.1% Pred. No. 1.2e-50;
Matches 281; Conservative 168; Mismatches 416; Indels 101; Gaps 34;

35 PTKESLSNKISLTGDTHTNL-----NCYLDNRLYLALILOKTPNEGAAVTITDYL 84
DB PLIDTILN---MTPYSHRATLFGVRDTONDIYLDHONSIESMFENSOGGALSCKS-L 102
47 85 SFPTDKEGIFYFAKNLTPESGALIGVSPSPVEIRDTIGPIVFEENNTCCRPFTSSNP 144
DB 103 AITNT-KNOILFLNFAIKRAGAM-YVNGN---FDLSENGSIIFFSGLN-----SFPN 150
145 AA--VVKIRREGCAIHAONLIVNHNDVVGPMKNSYVRGAISTANTFVYSENOGCLFM 202
151 ASNPADCTCTGGAVALCSKNVITSKNOGTATYFINNKAKSSGALIOAIIINIDNTGPTCLFF- 209
203 DNICIoTNTAG--KGAIVAGTSNPFESNCDLFFINNACCAGAI-FSPICSLTGKRG 259
DB 210 -----NNAAGTAGGALFANACR-IENNSQPIYFLNNGSGALIVHOCILTKNTGS 262
260 IVFNNRCFKNVETASBASDGAIKVTRLDVTGNRGRIFFESDNTKNGAIYAVVT 319
DB 263 VIFNNN--FAMEADISANHSSGAIYICIS-CSIKDNPGLAFDNTAARDGALICTOSLT 319
320 LVNCGPTFYFINNIANNKGAIYIDGTSNSKISADRHAIIFENENTVTVTNANGSTSNP 379
DB 320 IODSGPYEFTNN-OGTWGAIMLRQDGAICTLFAOOGDIIFNNRHEKDTSN--HVSVC 376
380 PRRAIIVASSSGEILLGASSONLIFYDPI---EVSNAVSVPFKKEDOTGSVVFSG 435
DB 377 TRNVSLTVGASG-----HSATFYDPILOKTYTIONS--IOKFNPNEHGLTIFSS 425
436 A-----TVNSADF--HORNLOTKTPAPLTLNSGFLCTEDHAQTLVNRFTOTGVVSLGNG 488
DB 426 AYPIDTSTSDDFISHFRN-----HIGLYNGTLALEDREKEMVUYFEDFGGLRLGSR 478
489 AVLSCKYKNGASNASITLKHIGLNSLISGAEIPLLMVEPTNNSNYTADTAATF 548
DB 549 SLSDVKLSLIDYGNSEPESTDLTHALSOPMLISSEASDNLRSDDMFESGLN-VPHYG 607
479 AVSTTDEOGSSSVGVININNLAINLPSTL-GNRVAPKIMIRPTGSSAPYSEDNNPII 537
DB 538 NLSG-PLSLDDENDLDPYDTADLQPIAEVPLLYLDVTAKHINTDNFREGGLNTQHYG 596
608 WQGLMTGMAKTODPEPASSATITDPOKANRFRHTLLTLWLPAGVYVSPKRSPLIANTL 667
DB 597 YQGVWSPYWIETITTSSTSS-----EDTVNLHRLQVGDMPRTGKYKVPENKGDIALSAF 651
668 W---GNMLLATESLKNSELTSPDHPFGITGGGIGMNVYODPRENHGPFMRSSGSYAG 724
DB 652 WQSHNLPLATLRYOTQOQIAPT-----ASGEATRLCFVHONSNDNAKGFMEATGYSLG 705
725 MIA--GOTHFESLKSOTYTKLINERYAKNNVSNKNSGCEMLPS-LOEGFLTKLVGLY 781
DB 706 TTSVASHSGVNFQSFNLYSHSDNSVASHITTVAILQINNPMLOERSTASSLA-Y 764
782 SYGHNCHHFTYQGEN--LTSQGTFRSOTMGCAVFFDLPKMPFGSTHLLTRPFLGALGIY 839
DB 765 SYSN---HHIKASYSQGIQTEGKCYSTTLGAALSCSISLO-WRRRLPHHTPFIOLAIVR 820
840 SSLHTEVGAYPRSFSTKPLINVLVIGVGSFMANATOPPOATVETLATOPVLYROEP 899
DB 821 SNOTAFQESGKAKRKFVHKPLNYLTVLGISQAMESKFLPTYWNLILAOPVLYQONP 880
900 GIATQLASKGIMFGSGSPSSRHAMSYKISOOTOPLSWLTLHFOYHGFYSSTFCNYING 959
DB 881 EVNVSLESSGSSWLLSGTLARNIAIAFKGRNQIIFPKLVFLDYGVSSTTHYIHA 940
960 EIALRF 965

DB 941 GTTFKF 946

RESULT 7
72075
polymorphic outer membrane protein e/f family - Chlamydia pneumoniae (strain CWL0
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence, revision 23-Apr-1999 #text, change 05-May-2000
C:Accession: C72075
R:Kalmn, S.; Matchell, W.; Marathe, R.; Lamm, C.; Fan, J.; Olinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MID:9920606
A:Accession: C72075
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-946 <ARN>
A:Cross-references: GB:AE001631; GB:AE001363; NID:94376750; PION:AAD18610.1; PID:9437
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: pmp_18

Query Match 18.6% Score 948.5; DB 2: Length 946;
Best Local Similarity 29.1% Pred. No. 1.9e-50;
Matches 281; Conservative 167; Mismatches 417; Indels 101; Gaps 34;

35 PTKESLSNKISLTGDTHTNL-----NCYLDNRLYLALILOKTPNEGAAVTITDYL 84
DB 47 PLIDTILN---MTPYSHRATLFGVRDTONDIYLDHONSIESMFENSOGGALSCKS-L 102
85 SFPTDKEGIFYFAKNLTPESGALIGVSPSPVEIRDTIGPIVFEENNTCCRPFTSSNP 144
DB 103 AITNT-KNOILFLNFAIKRAGAM-YVNGN---FDLSENGSIIFFSGLN-----SFPN 150
145 AA--VVKIRREGCAIHAONLIVNHNDVVGPMKNSYVRGAISTANTFVYSENOGCLFM 202
151 ASNPADCTCTGGAVALCSKNVITSKNOGTATYFINNKAKSSGALIOAIIINIDNTGPTCLFF- 209
203 DNICIoTNTAG--KGAIVAGTSNPFESNCDLFFINNACCAGAI-FSPICSLTGKRG 259
DB 210 -----NNAAGTAGGALFANACR-IENNSQPIYFLNNGSGALIVHOCILTKNTGS 262
260 IVFNNRCFKNVETASBASDGAIKVTRLDVTGNRGRIFFESDNTKNGAIYAVVT 319
DB 263 VIFNNN--FAMEADISANHSSGAIYICIS-CSIKDNPGLAFDNTAARDGALICTOSLT 319
320 LVNCGPTFYFINNIANNKGAIYIDGTSNSKISADRHAIIFENENTVTVTNANGSTSNP 379
DB 320 IODSGPYEFTNN-OGTWGAIMLRQDGAICTLFAOOGDIIFNNRHEKDTSN--HVSVC 376
380 PRRAIIVASSSGEILLGASSONLIFYDPI---EVSNAVSVPFKKEDOTGSVVFSG 435
DB 377 TRNVSLTVGASG-----HSATFYDPILOKTYTIONS--IOKFNPNEHGLTIFSS 425
436 A-----ATVNSADF--HORNLOTKTPAPLTLNSGFLCTEDHAQTLVNRFTOTGVVSLGNG 488
DB 426 TYPIDTSTSDDFISHFRN-----HIGLYNGTLALEDREKEMVUYFEDFGGLRLGSR 478
489 AVLSCKYKNGASNASITLKHIGLNSLISGAEIPLLMVEPTNNSNYTADTAATF 548
DB 479 AVSTTDEOGSSSVGVININNLAINLPSTL-GNRVAPKIMIRPTGSSAPYSEDNNPII 537
549 SLSDVKLSLIDYGNSEPESTDLTHALSOPMLISSEASDNLRSDDMFESGLN-VPHYG 607
DB 538 NLSG-PLSLDDENDLDPYDTADLQPIAEVPLLYLDVTAKHINTDNFREGGLNTQHYG 596
608 WQGLMTGMAKTODPEPASSATITDPOKANRFRHTLLTLWLPAGVYVSPKRSPLIANTL 667
DB 597 YQGVWSPYWIETITTSSTSS-----EDTVNLHRLQVGDMPRTGKYKVPENKGDIALSAF 651
668 W---GNMLLATESLKNSELTSPDHPFGITGGGIGMNVYODPRENHGPFMRSSGSYAG 724

```
Db 80 LSTNSSGAFV--ESMTSFTALDNADLFECCNNYCTHOGG--GGAINAGLISFKNNQNTL 135
Oy 128 IFENNFCRPTSSPNNAVKIRREGAIIHQNLYINHHVDVGFPMKNSYRGGAISTA 187
Db 136 FYNNTTICTOFTGVALKTERNR--GCALYGSSTIELINNHS--LNFINTNSGCMGCAVSTI 191
Oy 188 NTFVSENOGCFLEFMDNICIQ--TNT---AGKGAIIYAGTNSFESNNCDFEINNACC 241
Db 192 QNLVTKNTSGIYAFENNHTDHIPTPATILARGAIVGCGACEISHTNGPVENSNYCG 251
Oy 242 AGCAIIFS--PICSITGNKGNIVFYNNRCKNVTASSEASDGAIKVTRTLDVTGNRGRIE 300
Db 252 YGCAISTGCGCIEFRDNKKDLIFINNLSALGMHT--SAQNGAVISAGEFEGLNNKGPY 309
Oy 301 FSDNTKNGAIYAPVTLVDNGPTFYFINNIANKGAIYIDGT--SNKISADRHAIIE 359
Db 310 FENNASTIAGISCNNLNFQENGPYIFLNSA--LYGAFHLPASPAANYHTGSGDIIF 368
Oy 360 NENITVNTNAN-----GTSTSANPPRRNAITVASSSGEILLGAGSSONLIIFYDP 409
Db 369 NNNTLSTTGMAGLRKLEIPIGTNN-----NPITLS-----LGAKKOTRIIFYDL 415
Oy 410 IE-----VSNAVSVSFNKEADOTGSVFGCAIVNSADFHORNIQTYT-----PA 454
Db 416 FOMGLKRAPNPENSPHTVITNPDSFGAVFVSYKNI--SSDLQAHMIASTKHNOIKDS 474
Oy 455 PTLTNGFLCIEDHAO-----LTVNRETOTGGVYSLGNAVLSCKYKAGNASNASI 507
Db 475 PTLTNGFLTMSITNGABEFEPNGPLT---QESTSLALQODSLIYLYGK-----DASL 522
Oy 508 TLKHIGLNLSTLK--SGAEIPLWEP---TNSNNYATDRTATFSLSVKL-----SL 557
Db 523 TITHGILPGLINDGQTAPRIARVNPQDMQNTNSNQ--PVSTENAVTOKIPEFSGIYSL 581
Oy 558 IDYCNSPRESIDLTHALLSOPMLSTSEASDQNRSD---DMDFSLNPPHGMQGLW 614
Db 582 VQENESYVSDCLSRKGNAPLHLETINDAQLSNDMKNLTNTSLYSLPHYGOGLWTS 641
Oy 615 GNAKTODPPASSATITDOKANR-----638
Db 642 NMWTT--TPTVSLTNSSTETOTANNSTOEOKNSETFDSNSTTAKIPSTRASTGTTPLA 699
Oy 639 -----FHTLLTLWLPAGVSPKHSPLIANTLWGNMLLATESLKSNAELTPSDHPW 692
Db 700 TTDVYTRISLIVSVMPRIYIADPARGLIANNLVSSGNTLYLVS---LLPD--SWE 755
Oy 693 GTTGGGLGMVYQDPRENHPRFHMSSGY--SAGMIAGOTHTFSLFSGOTYTKLNERYAK 750
Db 756 ALQGSKATLFTKQOKRLDYHGYSSASKGYAISSQASGANGHKLFPFSSQSDMKKERTN 815
Oy 751 NNVSANVCQGEMLFSLDEGLITLYG--LYSYGDHNCHEFYTGENTSGTGRSOT 808
Db 816 NNTISSRYVL---SALCFEOPMEDRIALIGAANVNTGHTKTYNTF--GTCKFSGKGNHSTT 870
Oy 809 MCGAAYFDL--PMKPEGSTHILTPALGALGIYSSLHPTFVGAVPPRSFTKPTLINVLP 867
Db 871 LGGSLGCELRLDSMPQS--IMLPFIQALISRTREPASIOGQGLARLFLSKQHTAVVSP 928
Oy 868 IGVKGSFMNATORPQAWYELAYOPVLYROEPIATOLASKIHWGSGSPSRHAMSYK 927
Db 929 IGIKGYSSSKMPTVSCENEVAVOPTLYMKRPLNTLVLIKNNGSWETTNTPLAKHSF--YG 987
Oy 928 ISOOTOPLSMILHPOYHGFYSSSTFCNTYNGEIALRF 965
Db 988 RGSSSSLKFTLKLFPANTQAOVATSTVSHINAGCALVF 1025
```

```
RESULT 10
G71460
Probable outer membrane protein G - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
```

```
C:Accession: G71460
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A:Reference number: A71570; MUID:99000809
A:Accession: G71460
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1013 <ARN>
A:Cross-references: GB:AE001360; GB:AE001273; NID:q3329342; PIDN:AA68469.1; PID:q332
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: pmpG
```

```
Query Match 12.9% Score 655.5; DB 2; Length 1013;
Best Local Similarity 25.7%; Pred. No. 2,le-32;
Matches 261; Conservative 152; Mismatches 388; Indels 213; Gaps 43;
```

```
Oy 39 SLNNKISLTGDTNLTNLCYLDNLRYLALLOKTPNEGAAVTITDYLSPFDTOKEGIYFAK 98
Db 84 NLGSEFTVLGRHSILT--FENIR-----TSTNGALS-----DSANGSLF--- 121
Oy 99 NLTPESGAIAGTASPNSPVEIRDTIGPIFERNTCRREFTSNP--NAV----- 147
Db 122 --TIEGFKELFSNCSLLAVL---PAATTNNGSQTPTTSTPSNGTIYSKTDLLIN 174
Oy 148 -----KIR--EGGAIIHQNLYINHHVDVGFPMKNSYRGGAISTAFTVSENOG 198
Db 175 NEKFSYSLNVSGDGAIDAKSLTYGICKLFCQENTQAOAGCAOVVTSFSAAMENAP 234
Oy 199 FLFMDNICIOTNTAG--KGAIYA-----GTSNSFE-----SNNCDFEINNACCAG 243
Db 235 IAFIANV-----AGYRGGIIAAVQDQGVSSSTEDPVSFSSRTAVFEEDGNVAVRG 288
Oy 244 GAIFSPICSLTGRGNIVFYNN--RCEKN-----VERASSPASDGGAI 284
Db 289 GGILYS-----YGVNAFLNNGKTLFLNNVASPVYIAAEOPITNGASNTSDNYGDGGAI 340
Oy 285 KYTRRLDVTGN-----RGRIFESDNTKNGAIYAPVTLVDNGPTFYFINNIAN 334
Db 341 FCKNGQAQASNNSSGSVSFEDGEGVFFSSVVAAGKGAIIYAKKLSVANGCPQFLOINIAN 400
Oy 335 NKGCAIYIDGTSNKSISADRHAIIFNENT---VTNVTANNGSTISANPPRRNAITVAS 390
Db 401 D--GCAIYLGESGELSLADYGDIIIFDGNLKRATAENAADVNGTVSS---QAISMSG 454
Oy 391 SGEILLGAGSSONLIIFYDPLEVN-----AGVSYFKNKEADOTGSVYFSGATVNSADF 443
Db 455 GKITTRARAGHOILFNDRPEMANGNPOASSEPLKINDGEGYTDIYF--ANGNSTLY 512
Oy 444 HORNLOTKTPAPRLTSLNGFLCIEDHAOLTVNRETOTGGVYSLGNAVLSCKYKAGNAS-- 501
Db 513 QN-----VTIDQGRVLEKAKLSVNSISOTGSGLYMAGSLDLPVTTPQPOQR 562
Oy 502 ASNASITLKHIGLNLSTLKSGAEIPLWVEPTN--NSNNYATDRTATFSLSVKL--- 556
Db 563 AANQDLITLSLHLSLSSLLANNA---VTNPPTNPRAODSHPAIIGST--TAGSVTISGPI 617
Oy 557 LIDDYCNSPRESID-----LTHALLSOPMLSTSEASDQNLSDDDMFSGLVNPPH 606
Db 618 FFEDLDOTADAYRDWLGSNOK IDVLKIQLOTOP--SANAPSDTLT-----GNEMPK 667
Oy 607 GMOGLTWGAKTQDDEPASSATITDOKANRPHRTLLTLWLPAGVSPKHSPLIANT 666
Db 668 GYOGSKMLAN-----DPTNANNGPYTLAKATWTKTGTNPGGERASALVPSN 712
Oy 667 LMGNNMLATESLKSNAELTPSDHPW--GITGGGLGMVYQDPRENHPRFHMSSGY--SACM 725
Db 713 LMGS--IIDIRSAHSAIQASVDGRSYCGLVWSGVSNFFHYHDBRALGQGYRISGSGYLG-- 770
Oy 726 IACOTHTFSLKFSQOTYTKLNERYAKNNV--SSKNYSQCGMLFSLDEGLTLTLYVGLYS 783
```

Db 221 FSONLAKSGGALYSTDNCTITDN-FQVIFDGNASAMEAO-----AOGGALCTTTDKT 273
Oy 290 LDVTGNRGRIFFSDNITKNGAIVAPVTVLVNDGPTTFYINNANK-----GGAIIYIDG 344
Db 274 VLTGNGK-NLSTNNMTALTYGCAISGLAKSISAGGPTLFQSNISGSSAGCGGCAINIS 332
Oy 345 TSNKISADRAHAIIFENIIVTNANGSTSTSNAPPRNAITVASSGGEILGAGSSQNU 404
Db 333 ADELALSAISGDTFNNQVTN-----GSTST-----RNAINIDAKVTSIAAAGOSI 382
Oy 405 IFYDPIEVSNAVS-----FNKEADQTSVNSGATVNSAD-FHQRNLOTKRP 453
Db 383 YFDDPI--TNPTASTDTLNLNLADANSEIEGAIIVSGEKLSTPEALAAVSTIR 440
Oy 454 APLTSLNGFLIEDHQAOLTVNRFOTGGVVSGLNGAVLSCYKNGAGNSASNAITLKHIG 513
Db 441 OPAYIARGLVLRDGVTVTFKDLTOSPGSRILMDG-----GTLISAKKANLSINGLA 492
Oy 514 LMLSLILSGADEPLLWEPNNSNNYADTAATSLSDVKKSLIDDYNSPYESTDLTH 573
Db 493 VMLSSD-----DGTNKAALKEADKNISLSTG-TALIDTEG-SEVENHNLKS 538
Oy 574 ALSSOPMLISASDN-QLRSDMDPSGLNV--HYGNGLMTWGMKAKTODPEPASSAT 630
Db 539 A-STYPLELTTAGANGTTTIGALSTLLOEPETHYGYGNQOLSMAN-----ATSK 591
Oy 631 TPOKANRHRFTLLTWLPAGYVPSPKHRSPLIANTLMGNMLLATESLKNSEALTFSDHP 690
Db 592 GS-----INWTTGYIIPSEKSNLPLNSLMON-FIDINSLOLIEFKSSGP 638
Oy 691 F-----WGITGGGLMMVYODPRENHDPFHMRSRGYSAGMIAGOTHTFLKSOTYTKLNE 746
Db 639 FERELW--LSGIANFEYRDSMPTRGFRHISGVALGITATPAEDQTLFA--FCQLFA 693
Oy 747 RAKANNVSKN-----YSCQGMELFSLQELP-----LTKVLG----- 780
Db 694 R-DRNHITTKNGDITVYAGSLYFHHTGELFDI-ANFLMGKATRAPVWLSLSEIILPISPA 751
Oy 781 ---YSYGDHNCHEFTYOGENLTSGTFFRSOTMGAVFELPMKPFGSTHIL--TAPFLGA 835
Db 752 KRSYLTNDHMKTYT--DNSIIKSGMRNDAFCADGASLRF-VISVPLLKEVEFPVAV 808
Oy 836 LCIYSSLSHTEVGAVPRSFSTKTPILNLVPIGVKGSFNNATQRPQAVTEIAYOPVL- 894
Db 809 QYIYAHQODFYERYAEGRAFVN-KSELINVEIPIGVT---FERDSKSEKTYDULMIYID 864
Oy 895 -YRQEGITOLLASKGIVFGSGSPSRHMSKYSIQOTQPLSWLTLHFQYHGFYSSTF 953
Db 865 AVRNRKCTOTSLASDANMAYGTNLAROGFSYRANHFOVNPHEMLFGOF-AFEVRSS 923
Oy 954 CNVLNGEIALRF 965
Db 924 RMY-NTNLGSKF 934

RESULT 13

H81722
polymorphic membrane protein G family TC0263 (imported) - Chlamydia muridarum (strain N1
C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
C:Date: 31-Mar-2000 #sequence-revision 31-Mar-2000 #text-change 11-May-2000
C:Accession: H81722
R:Read, T.D.; Bunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
A:Reference number: AB1500; M010:20150255
A:Accession: H81722
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-987 <TEF>
A:Cross-references: GB:AE002293; GB:AE002160; NID:g7190298; PIDN:AAF39132.1; PID:g719030
A:Experimental source: strain N199 (Mopn)

C:Genetics:
A:Gene: TC0263

Query Match 12.2% Score 620.5; DB 2: Length 987;
Best Local Similarity 25.8% Pred No. 2.8e-30;
Matches 253; Conservative 153; Mismatches 404; Indels 171; Gaps 42;

Oy 39 SLSNKISLTGDTNHLNLCYLDNRLYLALLOKTPNEGAA-----VITDVLSPF 87
Db 82 NLGNFTINGRGHSLV--FENIR-----TSTGALSNHAPSGLEIVIAFDELISL 130
Oy 88 DTQKECTFAKULTPESGALIGASPNPTV-----EIRDITGPVIFENNTCCREPTSS 141
Db 131 NQNS-----LVSVPOTGGTT-TSPSNGTISRTDLVLRD-IRKVSFYNSLVS----- 177
Oy 142 NNAAVNIKREGAIIHQAOLYINHNHDVYGFPMKNFSYRVAAGIASTANTFVSENOCELF 201
Db 178 -----GDGALDAQSLMVGIEKLCTFOENVAOSDGCACQVTKTSVANGKVPISF 228
Oy 202 MDNICIQTNTAGKGAIVA-----GTSNSFE-----SNCCDLFFINNACAGATFSPIC 251
Db 229 LGNV-----AGKKGGGVAAVKQDQAGATDLDSVNFANNTAVFEENGSARIGGIIISD-- 281
Oy 252 SLTGNRGNIVFTNNR--CFKNVETA-----SSEASDGAII-KYTRLD 291
Db 282 -----GNISFLGNAKTVFLSNVASPIYVDPAAAGQPPADKMDNYDGGAIKNDTNIG 335
Oy 292 VYG--NRGRIFSDNITKNYGAIVAPVTVLVNDGPTTFYINNANKGAIYIDGTSNK 349
Db 336 EVSFKDEGVAFPSKNAACKGAIYAKKLLISDCGVQFLGNAAND-GGAIYLVDOCELS 394
Oy 350 ISADRAHAIIFENIIVTNANGSTSNAPPRRNATVASSGGEILGAGSSONLIFYDP 409
Db 395 LSAHDRDIIIFDGLKMAITOGAATVHDV-MVANSALSMATGGQITTLRKRKEGRILFNRP 453
Oy 410 IEVSNAG--VSSEFNKKEADOTGSVFS-GATVNSADPHORNLOTPAPPLTSLNGFLCI 465
Db 454 IEMANGQPIQTLTVNEGSGYTGDIYFAKGDV-----LYSIEISQGRILL 500
Oy 466 EDHQAOLTVNRFOTGGVVSGLNGAVLSCYKNGAGNSASNAITLKHIGLNLISLILK-SGA 524
Db 501 REQTKLLVNSLTQOTGGSVHMGSTLDF--AVTTPRANSMALTVNHFSLILNKNV 557
Oy 525 ELPILWEPNTN-----NSNNYADTA--TFSLSDVKLSIIDYNSPYESTDLTHALSSQ 578
Db 558 TNP-----PTNPVQVSPRAVIGNTAAGTVTISGP--IFEDDETRAYDNQMLGADQTI 610
Oy 579 PMLISASDONLRSDDMDPSGLNVPHYGMGLTWGMKAKTODPEPASSATITDPOKANR 638
Db 611 DVLQHLGANPRANAETDITLGNESSKYGYOGSMTLQW-----EPDPA-----NPPONS 660
Oy 639 FHRITLLTWLPAGYVPSPKHRSPLIANTLMGNMLLATESLKNSEALTFSDHPV-GITGG 697
Db 661 Y--MLASMTKTGYNPGPERVASLVNSLWGS--ILDVRSAHSAIQASIDGRAICRGIWIS 717
Oy 698 GLGMVYODPRENHDPFHMRSRGYSAGMIA--GQHTFSLKFSQTYTKLNERAKNNV-- 753
Db 718 GINSEFYHODALGQGYRHISGYSIGANSYFGSS-MFLGAFETETGR-----SKYVVC 771
Oy 754 SSKNYSQCEMELFSLQEGFLTKLVG-----LYSGDHNCHEFTYOGENLTSGTFRSQ 807
Db 772 RSNDDHCVSQVYLSLQALGSCLEFDADAVRASVYSGGNOHMKTSYFAEB--SNVMDNN 829
Oy 808 TNGGAVFELPMKPFGSTHIL--APFLGALGYSSLSHTEVGAVPRSFSTKTPILNL 865
Db 830 CYYGEGAGAPRLAASKLYLNLPRPVOERFVACHESSTEGDQAREKS-GHLMNLS 888
Oy 866 VPIGVKSFNNATQRPQAVTEIAYOPVLYROEGIATOLLASKGIMFGSGSPSRHMS 925
Db 889 IPIGVAFD-RCSSKHHPNKSFMGAIYICDAVRSISGETITLLSHKEITWTDAPHLAHGWM 947
Oy 926 YKISQOTQPLSWLTLHFQYHG 946

```

OY 441 ADPHOR--NLQTKTAPPLTSLNGFLCIEDHQAULTVNRFTOTGCVSLGCAVLSCCKNGA 498
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 426 ADEKAKADNFSTSLKOPRLASGTLAKGNVELDVNGFOTEGSTLL-----MQPGT 477
OY 499 GNSASASTLKHIGLNTSLKSGAEIPLWVEPTNNNNYADRAATFSLSDVLSLI 558
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 478 KUKADTEALSTLRKLVLDLAL-----EGNKSYSIEFAGANKTTLTIS-PLVRO 524
OY 559 DDYGNPSYESTDLTHALSSQPMLSISEASDNOLRSDMDPFGSLN-----PHYMGGL 611
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 525 DSSGNF-YESHNTNQAF-TQPLVFTNAT-----AASDIYDALTLSPVQTPREHYGQGH 578
OY 612 WTMGMAKTODPEPASSATTDPQKANRFRHTLLTWLPAQYVSPKHSRPLANTLGNM 671
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 579 WEATWA-----DTSTAKSGT-----MTWVTTGYNPNPERRASVDPDSLWASF 620
OY 672 LLATESLKSABETPDHDFW---GITGGGLMMVYQDPRENHPRGHMSSGYSAGMIAG 728
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 621 ---TDTRTLOQIMTSQANSISYQGRGLMAGSTANFPHKDKSGTQAFRHSYGYVGSAG 677
OY 729 ---QTHFSLEFSQTYTKLNERYAKNNVSSKNYSQ-----GEMLESLQ 769
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 678 DSENFESVAFQOLFCKDKDLFIYEN-TSHNTLASLYLQHRATLGLGMPSPGSIIDMLK 736
OY 770 EGFLLTKLVGLYSYGDNCHHFTT-----QGENLTSGQTFRSQTMGCAFFDLPMK-PRG 823
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 737 DIPLINLAOLSYSTKNDMDRTYSPKAGSWTNNSGALE---LGSLLALYLPKEAPFE 793
OY 824 STHILAPLGLALGITSLSHFTEVGAYPRSFSTKPLINVLVPIGVKSEFMNATQPOA 883
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 794 QGF---PFLKFOAVYSRQOENFESGAEARAP-DGDLVNCSTIPVGRLEKISEDEK-NN 848
OY 884 WTVELAVQPLVROEPGATOLLASKGIW 912
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 849 FEISLAVIGDYVRKKNRSLSLWVSGASW 877

RESULT 16
D72077
polymorphic outer membrane protein g family - Chlamydomophila pneumoniae (strain CML029)
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence, revision 23-Apr-1999 #text, change 05-May-2000
C:Accession: D72077
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: D72077
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-928 <ARN>
A:Cross-references: GB:AE001628; GB:AE001363; NID:g4376730; PIDN:AAD18593.1; PID:g437673
A:Experimental source: strain CML029
C:Genetics:
A:Gene: pmp_11

```

Query Match 11.7% Score 597; DB 2; Length 928;
 Best local similarity 26.1% Pred. No. 7.2e-29;
 Matches 233; Conservative 148; Mismatches 361; Indels 206; Gaps 47;

```

OY 36 TRESLSNKLSTLGD-----THNLJNCY---LDNLRYI-----IALIQTPNEGA 76
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 45 TPRTSATYSLTGDVFEYEPKGTPLSDSCFKOTDNLTFLGNGHSLTFGIDAGTHAGA 104
OY 77 AVITITDLSFTDQKRCIYAKKMLTPESGCAIGYASPNSTVEIRDTIGVITENNCCR 136
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 105 AASTT-----ANKNLTFSGFSLSLFSDSSPSTTV---TTGG----- 136
OY 137 PFTSSPNAVNRKREGAIAHONLVINHNHDVGFEMKNFSYVGAISTANTFVASENQ 196
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 137 -----GLSSAGGVNLENI---RKLIVAG---NSTADGAIKGA-SFLITGTS 178

```

```

OY 197 SCFLPNDNICIOTNTAGKGAITYAGTSNSESNNCL--PFINNACCAGAIFFP1CSLT 254
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 179 GDALFENN-----SSSTKGAI-ATTAGARIANNQYVFLSINASTSGAIDDEGTSIL 232
OY 255 GNRGNIVFVNNRCFKNVEVASSASDGAIVKTRTD-----VTCNGRIFFSDNITKNNG 310
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 233 SNKKFLYFEGN-----AAKTGGAIICNTRKASGSPELLISNNKTLIFASNAETSG 282
OY 311 GAIVAPVTLVDNGPTFYI-NNI--ANNKGAITYIGTSSKISADRAHAIIFENIVTV 367
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 283 GATHAKKLALSSGTFEFLRNVSATPRKGALSIDASGELSISATGNTIFRNLT-- 340
OY 368 TNANGTSTANPPRRNATIVVASSGELLGAGSSQNLIFYDPI--EVSNAVSVSFNKA 425
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 341 -----TTGSTDFPKRAINIGSNGKFTLEAKKNHTIFEDYDITSEGTSSDVLKINGSA 395
OY 426 ----DOTGSVVFSGCATVNSADFH-QRNLOTKTAPPLTSLNGFLCIEDHQAULTVNRFTQ-T 479
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 396 GALNPYQGTILFSGETLTADLKVADNLKSSFTQPVLSGGLLLOKGYTLBESTFSQBA 455
OY 480 GGVVSLGNGAVLSCYKNGAGNSASNTLKHIGLNLSTLKSAGAEIPLWVEPTNNNN 539
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 456 GSLGMDSGTTL-----STTAGSITITNGIIVDSL--GLKQPV-----S 493
OY 540 YTAATATFSLSDVLSLIDYGNSPYESTDLTHALSSQPMLSISEASDNOLRSDMDPS 599
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 494 LTKAGSANKYIVYSGKLNLIIDIEGNI-YES---HMFSDHOLFELKRTVDADVDTNVDIS 548
OY 600 GL-NVP-----HYMGQGLMTWGMAKTQDPEPASSATITDPQKANRFRHTLLTWLPAQY 652
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 549 SLIPVPAEDPNSTCYGFGQWNNVW-----TTDTATNTKAT-----ATWTKTGF 592
OY 653 VPSPKHSRPLIANTLWGNMLLATESLKSNAEL--TPSDHP--FWGITGGGLMMVYQDPR 708
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 593 VPSPERKSAVYCNLTWG-VFTDRLSLQOLVEIGATGMEHQGW---VSMNTFLKRTD 648
OY 709 FNNPGRFHMSSGYSAQMIAGOTHT-----FSLKFSQTYTKLNERY-AKNN----- 752
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 649 ENRKGFRHITSGGY---VIGSAHTPKDDLTFFAFCHLFAHDKCFIAHNNSRTYGGTLFF 705
OY 753 -----VSKKNYSQGMELTSLQ---EGF-----LLTLVGLYSYGDNCHHFTYQGENLTS 800
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 706 KHSHTLQPOVYLRILGAKRFSESAIEKPREIPLALOVVSFSIDNRMEETHYSLEP-S 763
OY 801 QGTFRSQTMGCAFFDLPMKPFGSTHILTAFLGALG---IYSSLSHTEVGAYPRSFST 857
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 764 EGSMSNECLAGIGLDLPF-VLSNPHPLKFTITPQKVENYVYQSIFESSSDGAGFSI 822
OY 858 KTVPLINVLVPIGVKSGFMNATORPQAMVELA--YQPLVYROEPGATOLLASKGIWFGS 915
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 823 GR-LNLNLSIPVGAH--FVQG-DIGDSYTYDLSGFVSDVYRNPOSTATLVMSDPMKIR 878
OY 916 GSPSSRHA 923
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 879 GCNLSROA 886

RESULT 17
F81591
polymorphic membrane protein G family CP0302 [imported] - Chlamydomophila pneumoniae (s
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence, revision 31-Mar-2000 #text, change 11-May-2000
C:Accession: F81591
R:Read, T.D.; Brumham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hicke
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis McpN and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255
A:Accession: F81591
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-949 <REA>
A:Cross-references: GB:AE002192; GB:AE002161; NID:g7189226; PIDN:AAF8159.1; PID:g718

```

A:Cross-references: GB:AE002192; GB:AE002161; NID:g7189226; PIDN:AAF8159.1; PID:g718


```

OY 369 NANGSTANPPRRNAITVASSSGEIL-IGAGSSONLITPDIEVS-NAGVSVSFNKEND 426
Db 351 SSSQITTT-----RNSINIGNTNAKIVOLRASOGNTIYEDPITTSITALSADLNLNCP 404
OY 427 Q-----TGSVVFSGATVNSADPHOR-NLOKTPAPLTLNSGFLICIEDHADLTVRFTQ 478
Db 405 DLGAPNATOGITVIFSGEKISEAEADNLKSTIOOPLTLAGQSLKSGVTLVAKSFQ 464
OY 479 TCGVAVSLGAVLSCYKNGAGNSASNA-STLTKHIGLNLSSILKSGAEIPLLWEPFNN 537
Db 465 SPGSLTLMND-----AGTTLETADGITTINNLVLNDSL-----KETKKA 502
OY 538 NNYTADTATESLSDVKLSLIDDYG-----NSPYESTDLTHALSQPMLSIASDN 589
Db 503 TLKATQASOTVTLTG-SLSLVDPSGCVYEDVSMNPNQVSCLTTLTADDPANHITDLAAD 561
OY 590 OLRSDDMDPSGLNVPHYGOGLTWGMMAKTQDPEPASATITDPOKANRPHRTLLTWLP 649
Db 562 PLEK-----NPIHWGYOGMMAISM-----QEDTAKSKNAAT-----LTWTK 597
OY 650 AGVPSPKHRSPLIANTLWGNML-----LATESLKSNAELTPSDHPFGITGGGLGM 702
Db 598 TGYNPNPERGRLVANTLNGSEFVDVRSIOQLVATKVRGQETR-----GIMCEGISN 650
OY 703 VYQDRENNHFGHMSGYSAGMIAGQTH-----FSLKFSQTYTKLMERRA-KNNVS 755
Db 651 FHKDSTKIKKGRHISAGY-----VVGATTTLASDNLITAFQCLFKDHFINKRASA 706
OY 756 KNYVSGOGEMLFSLQEGFLTLTKLVG-----LVSYGDNCHHFTYO---GEN-- 797
Db 707 YAAHLHLQHLATLSSSLRLRYLPDSESEOPVLFDQOISITIKNTKTYTQAPKRESS 766
OY 798 -----LTSQGTFRSQTMGAVFDDLPMKPGCSTHILTAAPLIGALGIYS 840
Db 767 YNDGCALELASSLPHTALSHGELFHA-----YF-----PFIKVASYI 804
OY 841 SLSHFTEVG-AIPRSSTKTPILNLVPIGVKSGFMNATQPOAMVELAYOVLVLRQEP 899
Db 805 HODSFERNRTTLVRSFDS-GDLINSVPIGITYFERFSRNER-ASYATVLYVADVYRKNP 862
OY 900 GIATOLASGKIFGSGSPSSRHA-----MSYKISQOTQPLSMLTLHFQ 943
Db 863 DCTTALLINTSKTGTGTLNLSRQAGIRAGICITAFASPLEVETNLSMEIR 912

RESULT 22
C72076
polymorphic outer membrane protein h family - Chlamydia pneumoniae (strain CML029)
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: G72076
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000. MIMD:99206606
A:Accession: G72076
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-978 <ARNP>
A:Cross-references: GB:AE001629; GB:AE001363; NID:94376734; PIDN:AMD16596.1; PID:9437673
A:Experimental source: strain CML029
C:Genetics:
A:Gene: pmp_14

```

```

Query Match 11.5% Score 583; DB 2; Length 978;
Best Local Similarity 26.1% Pred. NO.5.6e-28;
Matches 263; Conservative 137; Mismatches 411; Indels 196; Gaps 45;
OY 41 SNNISLTGDNHL--TNCYL-DNLRYLALLOKTPNEGAATITIDVLSFTDQKEGIYFA 97
Db 67 SSNLSLGGKLSLTFTSCQAPTNMVAL-----SAAETLV-FKNF-----SINFT 112

```

```

OY 98 KNLTPESGAIGVASPNPTVEIRDTIGVIFENN-----TCRPFSSNPNNAVKIREG 153
Db 113 GNOSTGLGLL-YGKD-----IVFOSIKDLITFTNRVAVSPASVTSKAP-AITTVTTG 164
OY 154 G-----AIAHONTLYINNHVDVGFMRKNEFSYRGCAISTANTFVVS--ENOSCELFMDN 204
Db 165 ASALQPTDLSLLEN--ISQSIKFEGLANF-----GSAISSPTAVAKFIMNTATMSFSHN 218
OY 205 ICITONTACKGCAIYAGTSNFSNCCDLFFINMAC-----AGALFS 248
Db 219 F-----TSSGGVIVYGSSLSLEFNNGCIIITFANSVNSLKGVTBSSGTALGSGAICI 273
OY 249 P-ICSLTGNRGNIYFNNRCFKNVTASSEASDGAIVTTRLDVTGRGRIFESDNT 306
Db 274 PTGTLELKNQCKTFYSN-----GTPNDAICAIET-CNIVGQCALDLSNTA 322
OY 307 KNYGCAIVAPVTLVDNCPYFIJNINANKGAIYI-----DQTSNKSISADRHAI 358
Db 323 ARNGGALCAKVLNIGRGPIEFSRNRA-EKGALIFGSPGDPAPKOTSTLTILASGDI 381
OY 359 PNEIVTNTNANGTSTSANPPRRNAITVASSSGEILLAGSSONLITFDPI-----E 411
Db 382 FQGNML-----NTRPGIRNAITVEAGGEIVLSAOGSSRLVFDYPTIHSLEPTS 430
OY 412 VSNAGVSFENKEADQTSVFSGATVNSAD-FHONLOTKTPAPLTLNSGFLICIEDHAO 470
Db 431 PSNKDITIAN--GASGVSVPFSKGLSTELLIPANTTILLGTGYKIASGEKITDNAV 487
OY 471 LTVNRF-TOTGCVSLGNGAVLSCYKNGAGNSASNA-STLKHIGLNLSSILKSGAEIPL 529
Db 488 VNLGFAITGSGGLTLGSGTLGLATPTGAPN--VDFTIGKLAFFPFLKND----- 539
OY 530 WNEPTNNSNNTYADTAATSLSDVKLSLIDYGNSEPYESTDLTHALSQPMLSIASDN 589
Db 540 FVASVYNAG--TKNVTILTGAL-----VDLE-----HDVTDLYDNVSLSPVAIPTA-- 584
OY 590 QLRSDMDPSGLNVP-----PHYGOGLTWGMMAKTQDPEPASATITDPOK 635
Db 585 -----FRGATVTKTGPQDELATPSHIGYOGKMSYMSRPLIPAPDGCGRGPPSS 636
OY 636 ANRHRITLITW-----LQAGVPSPKHRSPLIANTLWGNMLLATESLKSNA-----EU 684
Db 637 AN-----TLVAVMSDITLVRSTYILDERGEIYSNLSM-----ISFLGQARSDILOV 686
OY 685 TPSDHPFWGITGGGICGMVYQDRENNHFGHMSGYSAGMIAGQTH--TFSLKFSQTYT 742
Db 687 ILIDHGLSITAKALGAYEHTRPGCHEGFGRYGGOALSNMYTDHTTLGSLFQOLY 746
OY 743 KLNERAKNNVSSKNY--SCOGEMLSLQEGFLTLKVLGYLSYGDHCHHFTYQENL-T 799
Db 747 KTNANPYDSCSOMVLSLFFGQPIVTOKSEALISKKAATYISKHNLNTTYLRPDAPK 806
OY 800 SQCTFRSQTMGAVFDDLPMKPGCSTHILTAPE-----FLGALGIYSSLSHFTVGA 850
Db 807 SOGOMNNNSYVLSAE--HFLNMCLTLRPLAQAMDLSGFLISAERLEGWQSKFETGD 863
OY 851 YPRSFTKTPILNLVPIGVKSGFMNATQ-POAMVELAYOVLVLRQEGITATOLLASK 909
Db 864 LORSFS-RGKGYNVSLPIGCSQOMFPPKRAPSTLITIKLAKYKPDIVRVNHNITVVSNO 922
OY 910 GINFGSGSPSSRHMSYKISQOTQPLSMLTLHFQYHGFFSSFTFCN 956
Db 923 ESTSISGANLRHGLEVOI-HDVVDLTEDTQATLNTFTDCKNGFTTH 968

```

```

RESULT 23
B81593
polymorphic membrane protein h family CP0298 (imported) - Chlamydia pneumoniae (s
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: B81593
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salabe

```

QY 455 ---PILSLNGFLCIEDHQAOLTVNRFTQ-TGGVYSLGNGAVLSCYKNGAGNSASASTLTK 510
 DB 491 LNKLEGGTGLARNGATLVNHNFTQDEKSVYIMDAGTTLAT-TNGANN--TDGATITLN 547
 QY 511 HIGLNTSLKSGAEIPLWVEPTNNNNYADTAATFSLSDVKLSLIDDYGNSESTED 570
 DB 548 KLVINLDSL--DGTKAAYVAVOSTNGA-----LTIISGLGLVKNSSQDCDNGM----- 594
 QY 571 LTHALSSQPLMISSEASDNLRSQ-DMDFSGLVNPHYGMOGLMTGWAKTQDEPPASAT 629
 DB 595 FNRDLOQVPILELKATSNVTYTTDFSLGTNGCYQSPYQGTWEF---TID---TTTHT 647
 QY 630 ITPOKANRFRHRLTLTWLAGVPSPKHRSPLIANTLMGML-LATESLKSNAELPSD 688
 DB 648 VTG-----NMKKTGYLPHPERLAPLIPNSLMANVIDLRAVSOASADGEDVP 694
 QY 689 HPFNGITGGGLGMVYODPRENHPGFHMRSSGYSGAGMIAGOTHT-----FSLKFSQTY 741
 DB 695 GKOLSTG-----ITNFFHANNHG-DARSTRHMGGLINTYTRITPDALSLGFOLE 747
 QY 742 TKLNERAKNNVSKNYSQOGEMLFSLQEGFLTKLVGLSYGDHNGHFFYTOGENLT-- 799
 DB 748 TK-----SKDY-----LVG---HG-HSNVYFATVYSNITKS 774
 QY 800 -----SOGTFR-----SOTMGAVFFDLPKMPFGSTH 826
 DB 775 LFGSSRFSGTSTRVYTSRSNEKVKYTYKLPKRCMSNNCMWLGELGNLPITL--SSR 832
 QY 827 ILT-----APFLGALGIYSSLSHFTVEGAYPRSFSTKPTPLINVLPIGVKGSFMAATOPQ 882
 DB 833 ILNLKQIIPFVKAEVATVHGOIOENTPEGRIFG-HGHLVAVAVPVGRFG-KKSHNRPD 890
 QY 883 AMVELAYQPVLYROEPIATQLLASKGIWFGSPSSRHAMSYKISOOT 932
 DB 891 FYTIIVAVAPDVYRHNPDCTTLPLNGATWTISIGNNLTRSLVQASSHT 940
 RESULT 25
 C81593
 polyomphic membrane protein G family CP0299 (imported) - Chlamydomonas pneumoniae (stra
 C:Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
 C:Accession: C81593
 R:Read, T.D.; Brumham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis Mohn and Chlamydia pneumoniae AR39.
 A:Reference number: A81500; MUID:20150255
 A:Accession: C81593
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-995 <REA>
 A:Cross-references: GB:AE002191; GB:AF002161; NID:97189216; PID:AAF38156.1; PID:97189222
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: CP0299

Query Match 11.3% Score 575; DB 2; Length 995;
 Best Local Similarity 25.5%; Pred. No. 1.8e-27;
 Matches 273; Conservative 139; Mismatches 390; Indels 268; Gaps 50;

QY 1 MKKAFFFFLGNLSGLAREVPSRIFLMPNSVPPTKESLSNKI---SLTGDTHNLT-- 54
 DB 23 MKTSIRKFLIST--LAPCFASAFTEVIEIMPSENFQSSGKIFPYTTLSPRGTLICF 79
 QY 55 --NCYLDNLRIYLLAIOKT--PNEGAVTITDYLSPFTQKEGIFAKNL-TPESGGAIG 109
 DB 80 SGDLVIANLDNAISRTSSCSFNRAQALQILG-----KGVFSFLINRSSADGAAS 131
 QY 110 YASPSPIVEIRDTIG--PVIFENNTCCRPFTSSNPNAAVNRIRREGAIAHQNLTYINHH 167
 DB 132 SVITONPELCLSFSGFSQMTFDN---CESLTD--TSASNVIPIHSAIYATITPMLFTNN 186

QY 168 DVYGFMKNF-----YVKGAIANTPFVYSENGSCLEFMDN----- 204
 DB 187 DSILEFYNRSGAFGCAIRGISTILENT-----KSLFENGNGSISNGALTGSAIINLI 240
 QY 205 -----ICIQNTAG-KGCAIY-AGTSNFSFESNCDLFEFINNACCAGAIFFPICSLTGNR 257
 DB 241 NNSAPVIFSTNATGCIYGAIIYLLTGSMULTGSLGVLFPVNNSSSGAIIA-----N 292
 QY 258 GNIVFYNR--CEKNVETASSEAS-----DGAIAKV-----T 288
 DB 293 GNYTFSSNLDLEQN-WTASFPQNSLPAPTPPPPAVTPPLIGYGAIFCTPPATPPETGV 351
 QY 289 RLDVGNRGRIFESDNTIKNGAIVAPVTVLVNGCTVFNNIANNKGAITYDGTSSNS 348
 DB 352 SLTISGENS-VTFLENLASDEGALYCKKISIDSNKSTIFLGNTA--GKGAIAPESGEL 409
 QY 349 KISADRAIIFENIIVNTNANGSTSANPRRNATVVAASSGEILLGAGSSONLIYD 408
 DB 410 SLSANQDILEFNKL-----SITSGPTNNSIHFGKDAKFAILGATGCTTLTYFD 459
 QY 409 PI-----EVSNAQSVSFNKED--OTGSVFSGATVNSADFHQRLQKTPA----- 454
 DB 460 PITSDDLASASAAATVYVNPKASADGAVSGTIVSEGLTAT-----EATPANATST 512
 QY 455 ---PILSLNGFLCIEDHQAOLTVNRFTQ-TGGVYSLGNGAVLSCYKNGAGNSASASTLTK 510
 DB 511 LNKLEGGTGLARNGATLVNHNFTQDEKSVYIMDAGTTLAT-TNGANN--TDGATITLN 547
 QY 511 HIGLNTSLKSGAEIPLWVEPTNNNNYADTAATFSLSDVKLSLIDDYGNSESTED 570
 DB 570 KLVINLDSL--DGTKAAYVAVOSTNGA-----LTIISGLGLVKNSSQDCDNGM----- 594
 QY 571 LTHALSSQPLMISSEASDNLRSQ-DMDFSGLVNPHYGMOGLMTGWAKTQDEPPASAT 629
 DB 617 FNRDLOQVPILELKATSNVTYTTDFSLGTNGCYQSPYQGTWEF---TID---TTTHT 647
 QY 630 ITPOKANRFRHRLTLTWLAGVPSPKHRSPLIANTLMGML-LATESLKSNAELPSD 688
 DB 670 VTG-----NMKKTGYLPHPERLAPLIPNSLMANVIDLRAVSOASADGEDVP 694
 QY 689 HPFNGITGGGLGMVYODPRENHPGFHMRSSGYSGAGMIAGOTHT-----FSLKFSQTY 741
 DB 717 GKOLSTG-----ITNFFHANNHG-DARSTRHMGGLINTYTRITPDALSLGFOLE 747
 QY 742 TKLNERAKNNVSKNYSQOGEMLFSLQEGFLTKLVGLSYGDHNGHFFYTOGENLT-- 799
 DB 770 TK-----SKDY-----LVG---HG-HSNVYFATVYSNITKS 774
 QY 800 -----SOGTFR-----SOTMGAVFFDLPKMPFGSTH 826
 DB 797 LFGSSRFSGTSTRVYTSRSNEKVKYTYKLPKRCMSNNCMWLGELGNLPITL--SSR 832
 QY 827 ILT-----APFLGALGIYSSLSHFTVEGAYPRSFSTKPTPLINVLPIGVKGSFMAATOPQ 882
 DB 855 ILNLKQIIPFVKAEVATVHGOIOENTPEGRIFG-HGHLVAVAVPVGRFG-KKSHNRPD 912
 QY 883 AMVELAYQPVLYROEPIATQLLASKGIWFGSPSSRHAMSYKISOOT 932
 DB 913 FYTIIVAVAPDVYRHNPDCTTLPLNGATWTISIGNNLTRSLVQASSHT 962

Search completed: May 6, 2001, 19:18:48
 Job time: 543 sec


```

394 SerValIlePheSerSerGluArgLeuSerGluGluGlyThrProAs 410
      |||:
275 TTAGTATTATA.....CCTTC 259
410 PAsnLeuThrSer.....GlnLeuGlnInProIleG 421
      |||:
258 AAGCTTACTACCTTCATTATAATAATGTTCAATTGAAAAAGTCACTG 209
421 LuLeuLysSerGlyArgLeuValLeuLysAspArgAlaValLeuSerAla 437
      |||:
208 AAACCACTTCAGGCTCTTACTCCTTAACCA.....ACTGTT 171
438 ProSerLeuSerGlnAspProGlnAlaLeuLeuIleMetGluAlaGlyTh 454
      |||:
170 ACAAGCTCTTCTATTGAANA.....ATAGAAACTGTTTC 136
454 rSerLeuLysThrSerSerAspLeuLysLeuAlaThrLeuSerIleProL 471
      |||:
135 TTCTTTAAACACA.....ATCCTT 116
471 euHisSerLeuAspThrGluLysSerValThrIleHisAlaProAsnLeu 487
      |||:
115 TATTTATTTAGCAAGT..... 99
488 SerIleGlnLysIlePheLeuSerAsnSerGlyAspGluAsnPheTyrgI 504
      |||:
98 .....TCAGAAATACCAATTCTATTAGTGTGTAAT..... 69
504 uAsnValIleLeuLeuSerLysGluGlnAsnAsnIleProLeuLeuThrL 521
      |||:
68 .....CTTCATCATGTTTCAATAATAATTATTTATTACTA 32
521 euSerLysGluGlnSerHisLeuHis 529
      |||:
31 AAAGTAGAAAACTCTCTTCCAGCAC 6
seq_name: gb_es193:BF864732
seq_documentation_block:
LOCUS BF864732 741 bp mRNA EST 19-JAN-2001
DEFINITION 963054D01.x1 C. reinhardtii CC-1690, stress condition I, normalized
, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BF864732
VERSION BF864732.1 GI:12254876
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
REFERENCE Grossman, A., Davies, J., Federspiel, N., Harris, E., Hauser, C.,
AUTHORS Lefebvre, P., McDermott, J. P., Shrager, J., Silflow, C. and Stein, D.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Cellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 3
JOURNAL Unpublished (2000)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu
FEATURES
Source
1..741
location/Qualifiers
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, stress condition I,
normalized, Lambda Zap II"
/note="vector: pluescript II SK-; Site_1: EcoRI; Site_2:
XhoI; This library, constructed by John Davies and Jeffrey

```

```

McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr
, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into lambda zap II
(stratagene) in the EcoRI (5') and XhoI (3') sites.
pluescript II SK- plasmids were excised from the lambda
zap clones by superinfection with EXassist (Stratagene)
phase. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT 289 a 32 c 383 g 37 t
ORIGIN
alignment_scores:
Quality: 109.50 Length: 211
Ratio: 0.898 Gaps: 11
Percent Similarity: 57.820 Percent Identity: 29.384
alignment_block:
US-09-677-752-4 x BF864732/rev
Align seg 1/1 to reverse of: BF864732 from: 1 to: 741
339 LeuSerSerLeuGluAlaArgAsnGlyAspIleLeuPhe.....PheAs 353
|||||
702 CTCCTCTCTCTCTCTC.....CTCCCTACCCCTCTCTCTC 665
353 ProIleValGlnGlnSerSerLysGluSerProLeuProSerSerL 370
|||||
664 CTCGATCTCTCTCTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 615
370 euGlnAlaSerValThrSer...ProThrProAlaThrAlaSerProLeu 385
|||||
614 TCTCTCCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 565
386 ValIleGlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLe 402
|||||
564 CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 535
402 uSerGluGluGluLysThrProAspAsnLeuThrSerGlnLeuGlnInP 419
|||||
534 .....TCTCTCCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 498
419 rolleGluLeuLysSerGlyArgLeuValLeuLysAspArgAlaValLeu 435
|||||
497 CTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 448
436 Ser.....AlaProSerLeuSerGlnAspProGlnAlaLeuLeuI 449
|||||
447 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 398
449 eMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAla 465
|||||
397 TCTCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 348
466 ThrLeuSerIleProLeuHisSerLeuAspThrGluLysSerValThrI 482
|||||
347 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 307
482 euHisAlaProAsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyA 499
|||||
306 CTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 260
499 spGluAsnPheTyrgLysAsnValGluLeuLeuSerLysGluGlnAsnAsn 515
|||||
259 CTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 225
516 IlePro...LeuLeuThrLeuSerLysGluGlnSerHisLeuHisLeuPr 531
|||||
224 CTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 175

```


291 egiylgylalailealieglnserglyglyserleuSerlleuAlaag 308
 462 ATCGGGAAGTCTCG.....GGATTGAACTCTCGGCAAGCC 499
 308 lylgluglyserValleupheglnAsnAsnserGlnArgThrSerAspGln 324
 500 GAGAAAGCAGCGTACCTTTTGGAGGAGGATCTTTTACCGACTCTCC 549
 325 GlyleuValArgAsnAlaIleTyrLeuGlnLysAspAlaIle..... 338
 550 GGGGACACGTCAGGTCCTGGAGCTTGACAGGAGATCGGTGGAGTACAG 599
 339LeuSerSerLeuGlnAlaArgAsnGlyAspI 349
 600 CCGGAGGGAAGCGGCAACACACTGCTCTCCACTCGGCCAGCAAGACA 649
 349 lEleuphepAspProIleValGlnGlnSerSerSerLysGlnSerPro 365
 650 ACCGATGTCGAGGAGCGGAGGAGCTCTGGGACTTCTCTCGCAGAC 699
 366 LeuProSerSerLeuGlnAlaSerValThrSer.....ProThrPr 379
 700 CTGCGGAGAGTGTCTGCTCCCTCCACTCTCGAGTGGAGCGACGACTGA 749
 379 oAlaThrAla 382
 750 GAGAAAGGCC 759

seq_name: gb_est90:BF623271

seq_documentation_block: 1541 bp mRNA EST 18-DEC-2000
 LOCUS BF623271 HVSMBA0016E13f Hordeum vulgare seedling shoot EST library
 DEFINITION HVCDNA001 (Cold stress) Hordeum vulgare cDNA clone HVSMBA0016E13f,
 mRNA sequence.
 accession BF623271
 version BF623271.1 GI:11887101
 keywords EST.
 source barley.
 organism Hordeum vulgare
 Eukaryota: Viridiplantae; Embryophyta: Tracheophyta: Spermatophyta:
 Magnoliophyta: Liliopsida: Poales: Poaceae: Pooideae: Triticeae:
 Hordeum.
 1 (bases 1 to 1541)
 Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
 ,T., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
 ,T., Saski,C., Schwartzbeck,J., Simmons,J., Chol,D.W., Main,D. and
 Wood,T.
 Development of a genetically and physically anchored EST resource
 for barley genomics
 unpublished (2000)
 Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Seq primer: AATTAACTCTACTAAAGG
 High quality sequence start: 79
 High quality sequence stop: 967.
 location/Qualifiers
 1..1541
 /organism="Hordeum vulgare"
 /cultivar="Morex"
 /db_xref="taxon:4513"
 /clone="HVSMBA0016E13f"
 /clone_lib="Hordeum vulgare seedling shoot EST library
 HVCDNA001 (Cold stress)"
 /tissue_type="Seedling shoot"
 /lab_host="TJC121"
 /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 338 a 748 c 30 g 407 t 18 others
 ORIGIN
 Alignment_scores:
 Quality: 103.50 Length: 267
 Ratio: 0.877 Gaps: 12
 Percent Similarity: 44.195 Percent Identity: 24.345

Alignment_block:
 US-09-677-752-4 x BF623271 ..
 Align seg 1/1 to: BF623271 from: 1 to: 1541

351 phepAspProIleValGlnGlnSerSerSerLysGlnSerProLeuPr 367
 513 TTCTCAAACTCTACTACACACTACACCCCATACACACTCCACTCC 562
 367 oSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAlaSerP 384
 563 CTCACACACACACTCCACTCCCTCCCTCCCTCCCTCCCTCCCTCC 612
 384 roLeuValIleGlnThrSerAlaAsnArgSerValIlePheSer..... 398
 613 CATCTCTCTC.....TCACATTCATTCACATTCATTCACCTATNTC 656
 399SerGluArgLeuSerGlnI 405
 657 TCTTCCCTATATACACTCTTTTACATATCAACAACCATGGGGGAGAGA 706
 405 uGlnLysThrProAspAsnLeuThrSerGlnLeuGlnGlnProIleGln 422
 707 CCCCAACACTCT.....CATCCCTCTAACTC 735
 422 euLysSer.GlyArgLeuValLeuLysAspArgAlaValLeuSerAlaPr 438
 736 TCTTTCTCTCACAGCTCTCTCTCCACACACTCTCTTATCTCAACAC 785
 438 oSerLeuSerGlnAspProGlnAlaLeuLeuIleMetGlnAlaGlyThrS 455
 786 ANTA.....CCACCCCTCTTCACACACATACACACACT. 822
 455 erLeuLysThrSerSerAspLeuLysLeuAlaThrLeuSerIleProLeu 471
 823ACCTTCACCCCTTCCCTCCCTCTACTCTCCCTACCCCTT 864
 472HisSerLeuAs 475
 865 AACTCTCCAAAACAAATCAAAATGCTCTATACCCCACTCTCTCC 914
 475 pThrGlnLysSerValThrIleHisAlaProAsnLeuSerIleGlnLysI 492
 915 CTCACACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 964
 492 lEphLeuSerAsnSerGlyAspGlnAsnPhetYrGlnAsnValGlnLeu 508
 965 TTATCTCTCTCTTTATCATATACAC..... 990
 509 LeuSerLysGlnGlnAsnAsnIleProLeuLeuThr.....LeuSerLys 523
 991 CTATACACCAACTCTCTCATCTCTCTGATCACTCTCTCACTCTCATC 1040
 1041 TCCCTCTCTCTCTTATCTACTCTCTCTCTCAACCCCAACCCCTCTC 1090
 523 sGlnGlnSerHisLeuHis.....LeuProAspGlyAsnLeuS 536
 536 erSerHisPheGlyTyrGlnGlyAspTrpThrPheSerTrpLysAspSer 552
 1091 CCACCAT.....CACTCTTCC 1107
 553 AspGlnGlyHisSerLeuIleAlaAsnTrpThrPro..... 564
 1108 GACTATTCACACACACACCACTTACTACTCTCTCTCTCTCTCTCTAN 1157

and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGACC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGGACCGCATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 383 a 527 c 302 g 253 t 1 others
ORIGIN

alignment_scores:
Quality: 101.00 Length: 344
Ratio: 0.623 Gaps: 18
Percent Similarity: 47.093 Percent Identity: 23.837

alignment_block:
US-09-677-752-4 x BF570499 ..

Align seg 1/1 to: BF570499 from: 1 to: 1466

```

366 leuproSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAl 382
|||||
333 CTACCCAGTGCACACAG...TCTCTGTACGACCCAGCCTACTGGGCT 379
382 aserProLeu.....ValIleGlnThrSerAlaAsnArgSerValI 396
|||||
380 GCTTCACCTTACGCCCTATCCTGTATACACAGCACCCTTGGAAGCCCG 429
396 lephSerSerGluArgSerGluGluGluValThrProAspAsnLeu 412
|||||
430 TTATC.....AACTCTGTGATGAG 452
413 ThrSerGlnLeu.....GlnGlnProIleGluLeuLysSe 424
|||||
453 ACTTCCACGTGTACCCATCCTGTGTACAGCCACCTGTGTGACTGACTC 502
424 r.....GlyArgLeuValLeuLysAspArgAlaValLeuSerA 437
|||||
503 CACCGCACATTACGACCACTCTGGTAAGTCTTCACGCCCTGCTGG 552
437 laphSerLeu.SerGlnAspProGlnAlaLeuLeuIleMetGluAlaG 453
|||||
553 CACCACACTTGGAGCAAA...CGCTGGCCTTGTTCGATACAGTACGTG 599
453 yThrSerLeuLysThrSerSerAspLeuLysLeuAlaThrLeuSerIle 469
|||||
600 C.....CCAACTGGTACACCACTCAACCTGTGTAGTACGGGTATCC 640
470 ..ProLeuHisSerLeuAspThrGluLysSerValThrIleHisAlaPro 485
|||||
641 TTCCACAACACATCTAACGCCCAATACACATACATGAGGGGCCACA 690
486 .....AsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAsp 500
|||||
691 ATGGGGCGGCTTGGCATTTGGCCCATTTACAAACGGGGCCCAAGACAC 740
500 uAsnPheTyGluAsnValGluLeuLeuSerLysGluGlnAsnIleP 517
|||||
741 AAATTCACATCCAAACGCCCAACCTTATTCGTGGAGGAACACGGCAT 790
517 roleuLeuThrLeuSerLysGluGlnSerHisLeuHisLeuProAsp 533
|||||
791 CTGCACGGTCCACGGCGGTGGCCACCATTTCCACACACCT..... 834
534 AsnLeuSerSerHisPheGlyTyGlnGly.AspTrp.....ThrPhe 548
|||||
835 CGACATTGGACGGCTCAATGGACAGACGGCTTTCATGGACAGAAATTC 884
548 eTrpLysAspSerAspGluGlnHis.....SerLeuIleAla 560
|||||
885 CATGGCATACTAGGGTGGCCCAACACCAACCCCTTGGCGGTCCGACAG 934

```

```

561 AsnThrPheProLysAsnTyValProHisProGluArgGlnSerThrLe 577
|||||
935 AACACCCCCCGAGCGGTCTGCTCCCTTACACCATCACAAACAAAGCT 984
577 uValAlaAsnThrLeuTrpAsnThrTySerAspMetGlnAlaValGln 594
|||||
985 TTCACCAACA.....CATACAACTTTCATACCGTCATGGGGGT 1025
594 ermetIleAsnThrIleAlaHisGlyAlaTyLeuPheGlyTyTrp 610
|||||
1026 GCTGATTAGACCACTCGCAGCATGGCGGACAGCCTCATATGGGACA... 1072
611 GlySerAlaValSerAsnLeuPheTyAlaHisAspSerSerGlyLysP 627
1072 .....
627 oIleAspAsnTrpHisArgSerLeuGlyTyLeuPheGlyIleSer 644
|||||
1073 .....AACTACCCCGGACAGCAGCTTA.....ACAA 1097
644 hrHisSerLeuAspAspHisSerPheCysLeuAlaIleGlyGlnLeu 660
|||||
1098 CAGATGGCCAGACAGACACCATCGTCCGTCGCGGAGGCGAG..... 1141
661 GlyLysSerSerAspSerPheIleThrSerThrGluThrThrSerTy 677
1142 .....ACGCAACAATCACCCTATCG 1161
677 eAlaThrValGlnAlaGlnLeuAlaThr 686
1162 CACACACACCGGGCGCAACCGGCCACA 1189

```

seq_name: gb_est49:AW694070

seq_documentation_block:

LOCUS AW694070 665 bp mRNA EST 15-JUN-2000

DEFINITION NF072B11STP1092 Developing stem Medicago truncatula cDNA clone

ACCESSION NF072B11ST AW694070

VERSION AW694070.1 GI:7568807

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

Source

Location/Qualifiers

1..665

/organism="Medicago truncatula"

/db_xref="taxon:3880"

/clone_lib="NF072B11ST"

/clone_lib="Developing stem"

/tissue_type="stem"

/dev_stage="pooled developmental"

/note="Vector: Lambda Zap; Contains a mixture of

229	CGTGGGCGGAGCAACGTCGAGGTGTC...ATCGAGTCTCTGGGCGGC	345
163	PheThrAlaPheGluGluAsnSerSerLysGlyAsnGlyAlaIleI	179
346	TTCAACCCACATGAGAGAGGCCCTCCACCCACAGAGGAGTGGCGAAGA	395
179	nalaglnThrPheSerLeuSerArgAsnValSerProIleSerPheAla	196
396	GGTCCTGATCACCAGCCCGCTCGGCCACAGCCCGCCCATGTTGCTACGGTC	445
196	rgAsnArgAlaAspLeuAsnGlyAlaIleCysCysSerAsnLeuIle	212
446	TCAACCCACCATGCTGCTCAAGCGCCAGACAGGCTTCTTGTAACGCGCTCG	495
213	CysSerCylAsnValAsnProLeuPhePheThrCylAsnSerAlaThrAs	229
496	TGCACGACCAAC.....	507
229	nglGlyAlaIleCysCylSerAspLeuAsnThrSerGluLysGlyS	246
508TCCGCGGGCGCCCTCCACCAAGCATCAGACGACA	541
246	erLeuSerLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAla	262
542	AGTTCCGATGTCGCGCGCTGATGCACCCGTCACATCGACAGCGCC	591
263	LysGluLysGlyAlaIleTyrAlaIleLysIleMetValLeuArgTyrAs	279
592	ACCCAGAG.....ACCGTCA	608
279	nglYPro.....ValSerPheIleAsnAsnSerAlaLysIleGlyAla	294
609	CGCGCCCTCGGCGAAGCGCTGCGCGACGCGCGTGGCGCGCGCGACACA	658
294	IaIleAlaIleGlnSerGlyLysSerLeuSerIleLeuAlaGlyGlyGly	310
659	TCAATCCCTCGACGATCGCGCCCGCCAGGCGTC.....	693
311	SerValLeuPheGlnAsnAsnSerGlnArgThrSerAspGlnGlyLeuVa	327
693	693
327	IArgAsnAlaIleTyrLeuGluLysAspAlaIleLeuSerSerLeuGluA	344
694ACGCCGTCATCCCGACGTCACAG	718
344	IaArgAsnGlyAspIleLeuPhePheAspProIleValGlnGluSer..se	360
719	GCGCGCTGACTGCGATGCTTCGCGCGTCCCGACGCTCAAGTCTCGCTC	768
360	rSer.....LysGluSerProLeuProSerSerLeuGlnAlaAs	373
769	GTCGACCTGACTGCGCGATCAAGCCCGCCCGCTACGAGGAGTCTG	818
373	ervalThrSerProThrProAlaThrAlaSerProLeuValIleGlnThr	389
819	CAACCCGCTCGTGAGCGCCGCAACGCGGACGTCA.....AGGCG	859
390	SerAlaAsnArgSerValIlePheSerSerGluTyrLeuSerGluGluI	406
860	TCCTCGCGCTACACCGACGAGCGTCTGCTGCGAATCTGTCGCGGAG	909
406	uLysThrProAspAsnLeuThrSerGlnLeuGlnInPro...Ile..Glu	421
910	CCCATGCTGCTCATCTTGACGCCCAAGCGGGAATCGCCCTGACACAAAC	959
422	LeuLysSerGlyArgLeuValLeuLysAspArgAlaValLeu.....	435
960	CTTCGTCAGACTGGTCTCTGTTACGACACAGAGTGGGCGTACTTCGACAC	1009
436SerAlaProSerLeuSerGlnAspPro...GlnAlaLeuL	448

Email: rwing@clemson.edu
Seq primer: ATTACCCCTCACTAAAGG
High quality sequence stop: 1831

BASE COUNT	ORIGIN	
573 a	184 c	678 g
		351 t
		45 others

```
alignment_scores:
  Quality: 99.00
  Ratio: 0.497
  Percent Similarity: 44.619
  Length: 446
  Gaps: 18
  Percent Identity: 21.300
```

alignment_block:
US-09-677-752-4 x BE215233/review

Align seg 1/1 to reverse of: BE215233 from: 1 to: 1831

354 |||||ValIglInclUserSerSerylCylserProleuProserSeryl 370
1804 CCGGCTCCGCCCACTCCANTTCTTCACGTACCCTGCATCCATTAC 17555
370 uGlInalaserValThrSerProThrProAlarThalaserProLeuAlI 387
1754 TACTCTCTCTCTACTCCCACTCTCTCTTTCTTCTCTCTATTATTC 17050
387 legInThrSerLashsnrg..... 393
1704 CTCACACACACACATATCACACACACAGCAGCAGCTGGTTTAG 16555
394ServalllEpheserSerLuarfgleserGluGluGly 407
1654 AGCTTTTCTCTCTACTCTCTCTCTCTCTCTCTCTCTCTCTCT 16050
407 sThrProAspnslueThrSerGlnleugInGlnProIlectuleuylss 424
1604 ATGTCCAAATATATCCACT.....CCACTCCGGCTTCATCACACA 1564
424 ergIArgLeuValleuylsasparGalValleuSerAlaproSerleu 440
1563 CTCACCGGCTCGGTACC.....CTCTCTTCACCTAACCTC 1526
441 SerGlnasPro..... 444
1525 CCATCACTCCGCTGCTTGCTGTATTATACCGCTTATCTAGATT 1476
445GlnA 446
1475 GATTGTCCTCGCATATTGTTATCTTGACATTGTCATATCTTGATT 1426
446 lalenleuIllemelGlnlaGlyThrSerleuLys..... 457
1425 TCATTATCATATTAACATATTTCTCTTGCGGCATACCTCCCC 1376
458 ThrSer.SeraspLeuLysleualatnThrSerlleProleuhisertL 474
1375 ACGTCTTCACCTGCTTCACAGGCGCGCTCTCTCTCCCTCTTCTC 1326
474 euAspThrGlnLysSerValThrIleHisAlaProasnleu..... 487
1325 TCTCTTCACTCTCACTCAATCCAC...CACCGACGCACTTATCTTTAC 1279
488 SerIlegInLysllePheLeuSerAsnSerGlyAlapLunsnPheryrGI 504

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 19:25:51 ; Search time 27.13 Seconds
(without alignments)
3793.164 Million cell updates/sec

Title: US-09-677-752-4
Perfect score: 4533
Sequence: 1 MRPDHMFCCICAAILSSTA.....CVLRGSHSYSLDGLTTRF 878

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SPTREMBL_15:*

- 1: sp._archaea:*
- 2: sp._bacteria:*
- 3: sp._fungi:*
- 4: sp._human:*
- 5: sp._invertebrate:*
- 6: sp._mammal:*
- 7: sp._mhc:*
- 8: sp._organelle:*
- 9: sp._phage:*
- 10: sp._plant:*
- 11: sp._protist:*
- 12: sp._unclassified:*
- 13: sp._vertebrate:*
- 14: sp._virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4509	99.5	878	2	084882 chlamydia t
2	3543.5	78.2	867	2	09PL41 chlamydia m
3	947.5	20.9	928	2	092398 chlamydia p
4	919.5	20.3	930	2	09RB66 chlamydia p
5	915.5	20.2	930	2	092393 chlamydia p
6	896.5	19.8	926	2	P71135 chlamydophi
7	894.5	19.7	1276	2	09JRM2 chlamydia p
8	894	19.7	1407	2	092899 chlamydia p
9	883	19.5	928	2	09RB65 chlamydia p
10	853.5	18.8	914	2	086163 chlamydia p
11	845	18.6	928	2	086164 chlamydia p
12	845	18.6	949	2	09K299 chlamydia p
13	837.5	18.5	772	2	09RB71 chlamydia p
14	816.5	18.0	973	2	092896 chlamydia p
15	816.5	18.0	993	2	09K2A1 chlamydia p
16	790.5	17.4	846	2	P71133 chlamydophi
17	779	17.2	847	2	P71132 chlamydophi
18	776.5	17.1	839	2	P77792 chlamydophi
19	773	17.1	936	2	092898 chlamydia p

20	772	17.0	936	2	09J542
21	758	16.7	987	2	09PL45
22	758	16.7	1013	2	084879 chlamydia t
23	740	16.3	922	2	0929G5
24	739	16.3	922	2	09K1Y9
25	735	16.2	922	2	0924H9
26	708.5	15.6	841	2	0923A1
27	575.5	12.7	946	2	09RB60
28	574.5	12.7	946	2	0928B0
29	566	12.5	947	2	092813
30	566	12.5	947	2	09JSE2
31	562.5	12.4	649	2	P71134
32	557.5	12.3	978	2	09RB63
33	556.5	12.3	978	2	092895
34	546.5	12.1	712	2	09RB73
35	540.5	11.9	938	2	092883
36	531	11.7	934	2	092882
37	531	11.7	934	2	09JSE7
38	531	11.7	952	2	09K2A5
39	510.5	11.3	1520	2	09PLB0
40	503.5	11.1	1609	2	09RB58
41	503.5	11.1	1609	2	0926U5
42	493.5	10.9	986	2	09PY33
43	487	10.7	1016	2	084880
44	484.5	10.7	975	2	084417
45	482.5	10.6	976	2	09PL47

ALIGNMENTS

RESULT 1

084882 PRELIMINARY: PRT: 878 AA.

AC 084882:

DT 01-NOV-1998 (TREMblrel. 08, Created)

DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)

DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)

DE PUTATIVE OUTER MEMBRANE PROTEIN I.

GN PMPI.

OS Chlamydia trachomatis.

OC Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI_TaxID=813;

RN 111

RP SEQUENCE FROM N.A.

RC STRAIN=D/UM-3/CX;

RX MEDLINE=99000809; Pubmed=9784136;

RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L., RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., RA Davis R.W.;

RT "Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.";

RL Science 282:754-759(1998).;

DR EMBL: AF001361; AAC68472.1.;

SQ SEQUENCE 878 AA: 95592 MW; DFL1A31707BE48B CRC64;

Query Match 99.5%; Score 4509; DB 2; Length 878;
Best Local Similarity 99.5%; Pred. No. 9,9e+33;
Matches 874; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MRPDHMFCCICAAILSSTAVALFGODPLGETALLTKPNHYVCTFFPDCTMESLFPALCA 60
1 MRPDHMFCCICAAILSSTAVALFGODPLGETALLTKPNHYVCTFFPDCTMESLFPALCA 60
Db 1 MRPDHMFCCICAAILSSTAVALFGODPLGETALLTKPNHYVCTFFPDCTMESLFPALCA 60

QY 61 HASDDPLVYGNSYCWFSKLTHTDPKEALFKKGLSTONFRLSFTCCSSKSSPSI 120
61 HASDDPLVYGNSYCWFSKLTHTDPKEALFKKGLSTONFRLSFTCCSSKSSPSI 120
Db 61 HASDDPLVYGNSYCWFSKLTHTDPKEALFKKGLSTONFRLSFTCCSSKSSPSI 120

QY 121 IHKNGQLSRNNSMFCRNHAEAGSALISADAFSLQHNLYLFTAFENSSKNGAIOA 180
121 IHKNGQLSRNNSMFCRNHAEAGSALISADAFSLQHNLYLFTAFENSSKNGAIOA 180
Db 121 IHKNGQLSRNNSMFCRNHAEAGSALISADAFSLQHNLYLFTAFENSSKNGAIOA 180

Query	840	FRLTNORALHRLQTLNLCVLRQSHSISDLDGTYRF	878
DB	829	FRLTNORALHRLQTLNLCVLRQSHSISDLDGTYRF	867
RESULT	3		
ID	092398	PRELIMINARY:	PTT: 928 AA.
AC	092398		
DT	01-MAY-1999	(TrEMBLrel. 10, Created)	
DT	01-MAY-1999	(TrEMBLrel. 10, Last sequence update)	
DT	01-OCT-2000	(TrEMBLrel. 15, Last annotation update)	
DE	OUTER MEMBRANE PROTEIN PRECURSOR (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).		
GN	OMP10 OR PMP_9 OR CP0306.		
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).		
OC	Bacteria; Chlamydiales; Chlamydiales; Chlamydiales.		
OX	NCBI_TaxID=83558;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-CH1029.		
RC	MEDLINE=99206606; PubMed=10192388;		
RA	Kalman S., Mitchell W., Marathe R., Lammell C., Fan J., Hyman R.W.,		
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.;		
RT	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";		
RL	Nat. Genet. 21:385-389(1999).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-VRI310;		
RC	Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,		
RA	Christiansen G., Birkelund S.;		
RT	"Chlamydia proteins containing the GGAI-repeat belong to a subfamily		
RT	of autoantigenic pathogenicity factors.";		
RL	submitted (Jan-1999) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-J138;		
RA	Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,		
RA	Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;		
RT	"Comparison of whole genome sequences of Chlamydia pneumoniae J138		
RL	from Japan and CWL029 from USA.";		
RN	[4]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-AR39;		
RC	MEDLINE=20150255; PubMed=10684935.		
RA	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,		
RA	White O., Hickey E.R., Peterson J., Umayam L.A., Ulteberg T.,		
RA	Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,		
RA	Bowman C., Dodson R., Gwin M., Nelson W., Debroy R., Kolonay J.,		
RA	McClarty G., Salberg S.L., Eisen J., Fraser C.M.;		
RT	"Genome sequences of Chlamydia trachomatis Mopn and Chlamydia		
RL	pneumoniae AR39.";		
RL	Nucleic Acids Res. 28:1397-1406(2000).		
DR	EMBL: AE001628; AAD18591.1;		
DR	EMBL: AJ130343; CAB37069.1;		
DR	EMBL: AP002546; BA98655.1;		
DR	EMBL: AE002192; AAF38163.1;		
DR	TIGR: CP0306;		
KW	Signal.		
FT	SIGNAL.		
FT	CHAIN	1 27	POTENTIAL.
FT	28 928	OUTER MEMBRANE PROTEIN.	
FT	SEQUENCE	928 AA: 98332 MW: 58910A8F04F1219 CRC64;	
Query Match	20.9%:	Score 947.5, DB 2: Length 928:	
Best Local Similarity	31.6%:	Pred. No. 5, 2e-59:	
Matches 288;	Conservative 137;	Mismatches 336;	Indels 149; Gaps 32;
QY	85	TDPKALPFEKQDLSIONRFLSFIDCSKSSPSIIHOKNG-QISLRN-----	132
DB	52	TADGTVYVLDTGNTNAGSPALTALASCFTGTGLSFGHGQYCPDLLONIDAGACTTT	111

QY	133	-----NGSMSE-----CRINAECSGGAISADAFSLDHNTLFTAFPEENSSKGGAGAI	170
Db	112	NTANKLLISFGSEYLSLIQTTNATTTGTGAKISGTACISDSNT--SCYFGONFSNDNGAL	170
QY	179	QAQFFSLSRNVPISFARNRADLNGAICCSNLICSN-VNPLFTGNSATN-----	229
Db	171	QGSISLSLNNP-LTFAKKNKATQCGALYSNGGITINNTLSASFESNTAANGAIYTE	229
QY	230	-----GGAICCSIDLINTSEKSSLSLACNDELTFASNAKER	265
Db	230	ASSEFSSNKAISFINSVYATATATGCAIYCSS--TSAKPVLLTSDGELNFTGNTAITS	287
QY	286	GGAIYAKHMRYNGPVPSEFINS-----AKGGAIALQSGSGSILAGEGVYFONNS--	318
Db	288	GGAIYTDNLVYSSCGPILTFKNNKSLIDTAAPLGGAIALADSGSLSLAGDITFEENYAV	347
QY	319	-QRTSDQGLVNAATYL-EKDAILSLEARNGD-ILFDPPIVOESSKESPLPSLSQAVT	375
Db	348	KGASSQTTTNSINIGNTAKIYQLAASQNTIYFPDIT-----TSITAAIS	396
QY	376	SPTPATASPLVIQTSANRSVIFESSERLSEBEKT-PDNLQSQLOPIELASGRVLKORAV	434
Db	397	DALNINSGDLGNPAVOGTTIVFSEGEKLEAEADMLKSTIOQPLTLAAGQSLSKGYT	456
QY	435	LSAPLSDDPALLIMEGTSFLKTSDDLKATLSIRLHSDITDEKSVYIHPNLSIQIFL	494
Db	457	LVAKSFQSPSTLLMAGCTLTLEADGITINNLVLNVDSSLEKTKKATLKATQAS-OTVVL	515
QY	495	SMS-----GDENFEYENVELLSKEGDN--IPLTLRKEQ-SHLPLPDGNLS-----	537
Db	516	SGSLSLVPSGNYEDV-----SMNNQVYFSCLLTITADDPANHIITD--LAADPLEKPI	568
QY	538	HFGYQGDWTFPMKSDSECHSLIA--NMTPKNVYHPEROSTLVANTLMTNYSDMAVQSM	595
Db	569	HMGYGNMALSMOEDTATKSKAATILTYMTKTGYNNPERRGTLVANTLMSGFVYDRSIQL	628
QY	596	INTINHGAIYLFETMGSAVSLNLFYAHDSGKPIIDNNHHRSLGTYFGISTHSDHSCLA	655
Db	629	VATKYVROGOETRGICMEGISNF--HKDSTKINKGFPHISAGYVVGATTLASDNLITAA	686
QY	656	AGOLLGKSSDSFISTETTYTYIATVQAO-LAT-----SLMK-----ISAQACYN	698
Db	687	FCOLEGKRDRIPIKNNRASAALASHLOHLATLSPSLRLKTPSESEQVLRDAQSYI	746
QY	699	ESIHIELKTRYSFSEKFGSMHSAVAVSGEVCASIPYVS-NQSGLFSSFSIFSKLQGFSGT	757
Db	747	YSKNTMKTYTVOAPK-GSSWYNQGCALLETASSLPHTALSHEGFLHAYFPFIKEASYIH	805
QY	758	ODGEFE-SSGELRFSASAFKNISLPIQITTEKKSQKTRTYIYLGAYIDDKRDVESGP	816
Db	806	QDSFERMTTTLVRSQSDCLLNVSPVIGITFEERSRMRASRYEATVYVADVYKKNPDC	865
QY	817	VVLKNAVSMDAPMANLDSRA-----VWPERLTNGRALHRLDTLLNVSCVLGOSHSY	868
Db	866	TALLINNTSMKTTGTNLRSQAGIGRAGIFTAFS-----PNLEVTSLNLSMETINGSSRY	918
QY	869	SIDLCTYRF	978
Db	919	NADLOGKQF	928
RESULT	4		
Q9RB6			
ID	Q9RB6	PRELIMINARY:	PRT: 930 AA.
AC	Q9RB6		
DT	01-MAY-2000	(Tremblrel. 13, Created)	
DT	01-MAY-2000	(Tremblrel. 13, Last sequence update)	
DT	01-OCT-2000	(Tremblrel. 15, Last annotation update)	
DE	POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).		
DE	PMR_8 OR CP0107.		
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).		
OC	Bacteria: Chlamydiales: Chlamydiaceae: Chlamydia.		

```

Db 349 NTLTSTAPSTRNAIYLGSASAKITNLRAAOCOSIYFYDPIASNTTGASDVL----- 400
Oy 375 TSPATPASPVLVIOTISANSVIFSSERLS-EEKTPDNLTISLOQPIELKSGVLKDBA 433
Db 401 TINOPDSNSPL-----DYSGTIVFSGELSADEKKAADNTSILKOPLALASGLALKGNV 456
Oy 434 VLSAPLSODPOLLINEAGTSLKTSST-DLKLATLSLPLHSLDTEKSVTHAD-----N 486
Db 457 ELDVNFTQTEGSLTMQGTAKADTEALISLKLVLVDLSALEGNKSVSLETGANKTIT 516
Oy 487 LSIQKIFLSNGDENFYEENVELSKCONNIPLLITLSEOSHLPDGNLSS-----HF 539
Db 517 LPSPLVFODSSG--NFYESHTINOAFTOPLVFTTAASDIYI-DALLTSPVQTEPHY 573
Oy 540 GVGODMTSMKSDDEGHSLIANMTPKNYVPHPEROSTLVANTLTMTNTSDMOAVOSMINTI 599
Db 574 GVOGHAEATADTSTAKSGMTVTYTGYNPNPERRASVYVDSIMASTDTRTLOQIMTSQ 633
Oy 600 AHGAVLFGTWSAVSNLFYAHDSGKRPIDNMHRSLSGLYLGISTHSLDHSFCLAGOL 659
Db 634 ANSIYQORGLMAGSTANFPH-KDKSGTN-QAFPHKSYGYIVGSAEDFSSENFVAFCOL 691
Oy 660 LKGSOSFITSTETSTIATVQAQ-----LATSLMK-----ISAQACINES 700
Db 692 FCKDKDLFIVENTSHNYLASHLYLQHRAFLGGLPMPSPGSIITDMLKDLPLILMQLSYST 751
Oy 701 IHLKTKYRSFKESGEGSMHSAVSGVCASIP1-VSNGSLFSSFSIFSKLOGFSGTOD 759
Db 752 KNDMDRTYTSY-PEAQOSMTNNGSGALEGLSLALYLKPEPFQGYRPLKFAVARYRQ 810
Oy 760 GFEESGSEIRSFASFRNITSLPIGITEFKKSOKTRTYVYFLGAYIOLKRDVSGCVL 819
Db 811 NFRESGAEANAFDDGDLVNCISIPGILEKISEDEKKNFEISLAYIDGVYRKNPRSTRSL 870
Oy 820 LKNAVSDAPMANIDSRAYRFLTNORALH-RIOTLLNVSCVLRGSHS5LDLGTYYR 878
Db 871 MVSQASWTSJLCKNLARQAFLASAGSHLTLSPHVELSGEAAVEELGSAHY1NVDCGLRYRF 930

RESULT 6
ID P71135 PRELIMINARY: PRT: 926 AA.
AC P71135;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE PUTATIVE 98 KDA OUTER MEMBRANE PROTEIN.
OS Chlamydia phila psittaci.
OC Bacteria; Chlamydiales; Chlamydiaaceae; Chlamydia phila.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OVINE ABORTION S26/3;
RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U72499; AAB1818.1;
SQ SEQUENCE 926 AA: 98439 MW: 3E75E5EF594750F CRC64;

```

Query Match 19.88; Score 896.5; DB 2; Length 926;
 Best Local Similarity 30.3%; Pred. No. 2,2e-55;
 Matches 257; Conservative 145; Mismatches 341; Indels 105; Gaps 25;

```

Oy 98 LSIQNRFLSTDDSSKSSPSIIH-----QKNGQLSLRNGSGSPFRNIAEGSGAIS 151
Db 117 LTLTDFSKLSFKRCPSS-----SLVNTGKAMKSGGALMLANNASILFDONYSAEKGALIS 171
Oy 152 ADAPSLQNYLFTAFRENSKNGCAIOAQTFSS-LSRNVSPISFARNADLNGAICCS 209
Db 172 CKASFLTSKSSKEISFTTTSTAKKGALATATGAHLSNDQGTFRFGCNTAVNSGGAIVSEA 231
Oy 210 NLICSGNVNPLFTGNSATN-----GGAICCSIDLNTSEKGSLSIACNOETLFPASNSAKE 264

```

```

Db 232 SMTIAGN-NHVAFSNNNAVSGSSDGGCGALHC---SKTGSAPLITIDKNKVLFEENTSSA 287
Oy 265 KGAIAVKAHMVLYRNGPVSEFINN---SAKIGGAIAIOSGGLSILAGESVLFPONNSOR 320
Db 288 KGAIAITDKLILITSGGPTAFINNKVTHATPKGAGIAGIANGECSLTAHEGDIITFDNNLMA 347
Oy 321 TSDQGLV-RNATYLEKDALISLEARNG-DILFFDPIVDESSSKESPFLSSQASTSTPT 378
Db 348 TQDNATIKRINAIEGNGKRVNLAASGKTISEYDPIVEGNAAADLLLNKAEGKXT--- 404
Oy 379 PATASPLVIOTSANRSVIFSSERLSEEB-KTPNLTLSOLOPPIELKSGRLVLDRAVLSA 437
Db 405 -----YNGRIIFSGEKLTEQAAVADNKLTTPTPTTLTLAGELVLRSGVEVA 452
Oy 438 PSLSDPOLLINEAGTSLKT-SSDKLATLSLPLHSLDTEKSVTHA-----PNLSIQKI 492
Db 453 KTVYQTAGSLILMDAGTKLSAKTEDATLTNLAINPTLDGKFAVVDVAAGKRVTLTSGA 512
Oy 493 FLNSGDENFEN-----VELSKCONNIPLLITLSEOSHLPDGNLSSHFCY 541
Db 513 IGYIDPTGKFEYENHKLNDTLALGGLSGKGS-----VTTTNPVSHV--VGVAETHYGY 564
Oy 542 QGDMTSM-KDSD---EGHSLIANMTPKNYVPHPEROSTLVANTLTMTNTSDMOAVOSM-- 595
Db 565 QGMWVSVMVADNNSDKRTQTAIFTWKTYVPPERRAPVLYNSLMSFTDLASIQDYLE 624
Oy 596 -INTIAGGAVLFGTWSAVSNLFYAHDSGKRPIDNMHRSLSGLYLGISTHSLDHSFC 653
Db 625 RVSYDSILETRR---GLWVSGIGNFF--HKDRNENKRFRIHSSGYVLGATNTTSREDSLS 679
Oy 654 LAAGQLKGSOSDFITSTETSTIATVQAOLATSL-----MKI 691
Db 680 VAFQQLFAKDKDYLKSNKAAVYAGSVYQHVSGFDDLTBLFNGPNTCGSGFKEPIPL 739
Oy 692 SAQACYNESIHELKTKYRSFKESGEGSMHSAVSGVCASIP1VYNSGSLFSSFSIFSKL 751
Db 740 DAQITTYCHTANMTTSTYD-PEYKSGMGNDTLGLTSTVPIPVSSSIFDSTAPPAKL 798
Oy 752 QGESGTOGFEESGSEIRSFASFRNITSLPIGITEFKKSOKTRTYVYFLGAYIOLKRD 811
Db 799 QVYVAHODDEKPTTEGRVSESSDLNVSPDIGIKFEKLSYGERSAVDLTLMYIPQYRH 858
Oy 812 VESGPVVLKNAVSDAPMANIDSRAYRFLTNORALH-RIOTLLNVSCVLRGSHS5YSL 870
Db 859 NPSGMTLAINVSWLTAATNLARQAFIVRAGNHIALTSGVEMFSOPFELRSSRMYNV 918
Oy 871 DLGTYRF 878
Db 919 DLGAKVAF 926

```

```

RESULT 7
ID Q9JRW2 PRELIMINARY: PRT: 1276 AA.
AC Q9JRW2;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G/I FAMILY (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).
GN PMP_6 OR CP0309.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaaceae; Chlamydia phila.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RA MEDLINE=20330349; PubMed=10871362;
RA Shiral M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K., Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CW029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).

```

QY 521 LSKESQSHLHPDGN-----LSHFEGYQDWTFSWKDSEGH-SLIANTPTKNVYHPER 573
 DB 1024 LSTSGTAVLDLDFNPIPSMAAPDYQSGSMTLVPKVGAGKVTILVAEWAQALGYPKPEL 1083
 QY 574 QSTLVANTLVNTYSDMQAVQSMINTIAHGAVLFGTWGSAVSLFPAHDSGCPIDNMHH 633
 DB 1084 RALVPLNSLWNVNINHISIOEATATMSDAPSHPIGIMIGIGNAF--HODKOKENAGFRL 1141
 QY 634 RSLGKYFGISTSLDHSFCLAGOLLGSGDSFISTETSTETTYATVQAO-----LATS 687
 DB 1142 ISRGYIVGSSMTTPQETRAVAFSOLFSGSKDYVSDISQYAGSLCAOSSVITLHSS 1201
 QY 688 LMK-----ISAOACYNESIHELKTRYSFSGEFGSWHSAVSEVCA 730
 DB 1202 LRRHVLKVLPELPGETPLVLRHQVSYGRNHNMTTKLAN-NTQKSDMDSHFAVEVG 1260
 QY 731 SLPIVNSGSLFSSSIFSKLQSGTQDGFESSGEIRSFSSFRNISLPIGITEKK 790
 DB 1261 SLPVDLNTYRL-TSYSPRYKLVQVSVNOKGFQEVADPRIFDASHLVNWSIPMGLTFKHE 1319
 QY 791 SOKTRTYFFLAGYIQLDRVDSGPVLLKNVSWDAPMANLDSRAYMRLTNO--RALH 849
 DB 1320 SAKPPSALLTLGAYAVDAYRD-HPHCLTSLTNGTSMSTFATNLSDRONFREAASGHLKLH 1378
 QY 850 RLQTLNVCVLRGQSHSYSLDLGITYRF 878
 DB 1379 GLDCFASGCELRSSRSRYNANGCTRYSF 1407
 RESULT 9
 Q9RB65 PRELIMINARY: PRT; 928 AA.
 ID Q9RB65: Q9RB64: Q9SBP2:
 AC 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (OUTER MEMBRANE PROTEIN 5
 DE PRECURSOR) (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).
 GN PMP.10 OR OMP5 OR CP0303.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83358;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-J138:
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kunara S., Nakazawa T.,
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CWL029 from USA."
 RL Nucleic Acids Res. 28:2311-2314(2000).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-VRIJ10:
 RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
 RA Christensen G., Birke Lund S.,
 RT "Chlamydia proteins containing the GCAT-repeat belong to a subfamily
 RT of autoexporting pathogenicity factors."
 RN Submitted (JAN-1993) to the EMBL/Genbank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN-AR39:
 RA MEDLINE-20150255: Pubmed-10684935:
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Unayam L.A., Uterback T.,
 RA Berry K., Bess S., Linher K., Weidman J., Khoriti H., Colaven B.,
 RA Bowen C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonen J.,
 RA McClirly G., Salzberg S.L., Eisen J., Fraser C.M.,
 RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
 RT pneumoniae AR39."
 RL Nucleic Acids Res. 28:1397-1406(2000).
 DR EMBL: AP002346; BA98657.1;
 DR EMBL: AJ133034; CAB37071.1; -;
 DR

DR EMBL: AE002192; AAF38160.1; -;
 DR TIGR: CP0303; -;
 KW Signal.
 SQ SEQUENCE 928 AA: 97229 MW: 059005206A1DD0E1 CRC64:
 Query Match 19.5% Score 883: DB 2: Length 928;
 Best Local Similarity 30.6%: Pred. No. 2.1e-54;
 Matches 288; Conservative 145; Mismatches 401; Indels 108; Gaps 31;
 QY 12 CAALISTAVLFCQDPL--GETALLTKPNHVC-----TFEEDCTESTL--FPALCAHAS 63
 DB 20 CSTVFATAIENICPSDFDSTNTGTYTPKNTTGTIDYITLIGITLONLDSALTKGCF 79
 QY 64 QD--DPLVYLGNSYCWVFSKRLHTDPKEALFKKGLDSIONFRLSFTDCSKE--SSPS 119
 DB 80 SDTTELSFAGKCY--LSPLNIKSSAEG-----AALSVTDKNLSLTFGSSILFLAAPS 132
 QY 120 IT-----HOKNQLSLRNNGSMFPCNNHAEAGGGAISADAFSLQHYLFTAEENS 170
 DB 133 SVITTPSGKCAVCGCDLTFDNNGTILFKODYCEENGGAISTNLKLNKSTGSISEGNK 192
 QY 171 SKG---NGCAIQAO--TPESLRNVSPISFARNRADLNGAI--CCSNLCSGNVPLPFTGN 225
 DB 193 SSATGKKGAICTGTVDTINNTAPTLFSNNIAEAGAGINSTGCTIGNTS-LVFSFN 251
 QY 226 SAT-----NGCAICISIDLNTSEKGSLSLACQETLPFASNSAKEKGAIAKHWLYRY--N 279
 DB 252 SVTATANGCAL-----SGDADVTISGNQSVTFSGNQAVANGAIAKYLTLASGCG 303
 QY 280 GPVSFLNN-----SAKIGAIQSGSLSILAGEGVLFQNNNS--ORTSDGLVRNAYL 333
 DB 304 GGISFSNNIVQGTAGNGAISTILAEGCSLSAEGATIFNGAIAVATTPQTKRNSIDI 363
 QY 334 EKALISLSEARNG-DILFPDPIVQESSKESPLPSLSQASVSPPTATSPVLIOTSAN 392
 DB 364 GSTAKTILNLRAISGHSTIFFDPTTANTADSTDLNKNKADAGN-----STDYS 412
 QY 393 RSVLFSSEIRLSEEE-KTPDNLTSQLQPIELKSGRLVLRDAVLSAPSLSQDQALLIME 451
 DB 413 GSVFSGEKLSDEBAKADNLSTLKPVTLAGNLVLKRGVTLDTGFGTQAGSSVIMD 472
 QY 452 AGTSLKTS--DLKLATLSTPLHSLDTEKSVTIAP-----NLSIQKIFLSNGDEFTYE 504
 DB 473 AGTTLKASTEEVTLTGTSIPVDSLGECKKVIAASAASKNVALSGPILLDNOG--NAYE 530
 QY 505 NVELLSEKON-----NPLPLLSKQSHLHPDGNLSHFEGYQDWTFSWKD-----SDEGH 556
 DB 531 NHD-LGKTQDFSEVOLALGTATTDDVPAVPYATPTHTYGTGWTGNTVDDTASTPKTK 589
 QY 557 SLIANTPKNVYHPFEROSTLVANTLVNTYSDMQAVQSMINTIAHGAVLFGTWGSAVSN 616
 DB 590 TATLAWNTGYLNPBEROGLVLPNSLWGSFSDIQAIGYIERSALUTLCSDRGFMAACVAN 649
 QY 617 LEVAHDSGKPIDMWHRSUGLYFGISTHSDHSFCLAGOLLGSGDSFISTETTSY 676
 DB 650 -FLDKDKKCEK-RKYRHRKSGGVAIGCAAOCTCEMLISFAPCGLFGSKDKPLVKKNHDTY 707
 QY 677 IATVQAOALATSL-----MKISAACYNESIHELKTRYSFSGEFG 717
 DB 708 AGAFYIOHITCEGCGFICGLDLKPLGSMHKLPLVLEGOLAYSHVNDLKITYTAI-PEVKG 766
 QY 718 SWSVAASGEVCAISPIVNSGSLFSSSIFSKLQSGTQDGFESSGEIRSFSSASFR 777
 DB 767 SMCNNATNMWMLGASSHSYPEVLRHCFDTYAPYIKLNLTYIHODPSSEKGTGRSPDOSNLF 826
 QY 778 NISLPIITTEKKSOKRTTYTYFFGAVIQLDKRVESGPVLLKNVSWDAPMANLDSRA 837
 DB 827 NLSLPIVCKFEKESDCNDFSVDLTLVSPVDLIRNDPKCTTALVSGASWETVANNLARO 886
 QY 838 YMFRLTNGRALHRL-OTLVNVCVLRGQSHSYSLDLGITYRF 878
 DB 887 LQVAGSHYAFSPMFEVLGQFVEFVRGSSRIYNDLQKGFQF 928

KW Signal 1 17 POTENTIAL.
FT SIGNAL 18 928 OUTER MEMBRANE PROTEIN 4.
CHAIN 18 928
SEQUENCE 928 AA: 98903 MW: 788BCDD62C911402 CRC64;

Query Match 18.6% Score 845; DB 2; Length 928;
Best Local Similarity 28.8%; Pred. No. 1e-51;
Matches 247; Conservative 148; Mismatches 345; Indels 118; Gaps 26;

97 DLSIQNREFLFTDCSSKESPPS-IIHOKNGOLS-----LRNKGSMFCNNHAGSGCA 149
113 NLTFSGSLISF-----DSSPTVTYTGQGLSSAGGVNLEIRKLVAGNFTADGCA 166
150 ISADAFSLQHNLYLFTAFENSSKNGGALQAOQFS-LSRNVSPISFARNRADLNGAIC 208
167 IKGASFLITGSGDALFNSNSSSTKGCALATTAGARIANNNGYVPLSNIASTSGAID 226
209 SNLISGVNPLFTTNGSA-TNGAICISDLNTEKGS--LSLACNOETLPASNAKEK 265
227 EGTSLISNNKFLYFEGNAKTGGAIC-----NTKASGSPELIISNNKTLIFASVAVTS 281
266 GGAIYAKHMYLRNGPVSFINNSAKI-----GGATAIOSGSLISLAGGSVLFQNSQRT 321
282 GGAIAHAKKLALSSGGFTFLRNWSSATPKGALSIDASGELISAETGNIIFVANNITLT 341
322 --SDQGLRNALYLEKDALISLSE-ARNGDILFDPPIVOESSKESPLSSLOAVTSPT 378
342 TGSITDTPRRNAINIGSNCKFTELRAKKNHTIFFYDPTISEGSSDYLVKINNGSAGALNPY 401
379 PATASPLVYIOTSANRSVIFSSERLSEEE-KTPDNLTSQLQPIELKSGRLVLDRAVISA 437
402 QGT-----ILFSGETLTADELKADNKLKSSFTQPVSLSGGLLQKCVTLSS 448
438 PSLSDPPALLIMEAGTSKLTSS-DLKLATLSIPLHSLDTEKSVIIRHAPNLSIOKIFLSN 496
449 TSFSGEAGSLIGMOSGTLTSTAGSITITNLGINVDSLGLKQPVSLAKKAS-NKVIYSG 507
497 --SGDENFEYENVELLSKEO-----NNIPLLTSLKESQSHLHPDGNLSS 537
508 KLNLDIDEGNIYES-HMFSHDLFLSKLITVDADVDNVDISL-----IYVPAEDRPS 560
538 HFGYOGDMTFWSKDSDEGHS--LIANWTPKNVYVPEROSTLVANTLWNTYSDMQAVOSM 595
561 EYFGOGQMVNMTTDTATNTEKATATWTGTGFPVSPERKSALVNCNTLWGVFDIRSLQDL 620
596 INTIAHGCAVLEFGTWSAVSNLFYAHDSGKPIDMWHHSLSGLYLGISTHSLDHSFCLA 655
621 VEIGATGMEHKOGFWVSSMTN--FLHKTGDENRKGFRHTSGGYIGGSAHTPKDDLFTFA 678
656 AGOLLGKSSDSFTSTETSYIATV-----QAOLATSLMK----- 690
679 FCHLARDKDCGAIANNSTGTGTFEFKHSHTLOPQNTYLRGRAKFSEAIKFPREIRPL 738
691 -ISAQACYNESIHLEKTKRFSKSGFWSHSAVAVSGEYCAP-IVSNGSLFSSFSIF 748
729 ALDVOVFSHSDNRMEHTYSL-PESEGSMSNECIAGIGLDLPVLSNPHLFTFTFIPO 797
749 SKLQFGSGIDGDESSSGIRSFSSAFRNISLPIGITEKKSOKTRITYYVLGAYIOL 808
798 MKVENYVYVONSFFSSSDGRGFSIGRLNLISIYVGAKVQVQDIDGSYTYDLSGFFVSIV 857
809 KRDVSGPVLLKNAVMDPAMNLDRAVYFRLTNQALHRLQTLNLSV-----V 860
856 YANNOSTITVLVMSDPMKIRGNGLSROAFLRGSN-----NYVYNSNCELFGHYAME 910
861 LRQSHSYSLDLGTYRF 878
911 LRSSSNYVNDVGTCLRF 928

ID 09K299 PRELIMINARY; PRT: 949 AA.
AC 09K299:
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE POLYMORPHIC MEMBRANE PROTEIN G FAMILY.
CN CP0302.
OS Chlamydia pneumoniae (Chlamydia phila pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia phila.
OX NCBI_Taxid=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39.
RX MEDLINE=20150255; Pubmed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Umeyam L.A., Uterback T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven J.,
RA Bowman C., Dodson R., Gwin M., Nelson W., Debey R., Kolonay J.,
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Morn and Chlamydia
RL pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL: AE002192; AAF38159.1; "
DR TIGR: CP0302; "
SQ SEQUENCE 949 AA: 101357 MW: A00B09E16C699BE3 CRC64;
Query Match 18.6% Score 845; DB 2; Length 949;
Best Local Similarity 28.8%; Pred. No. 1.e-51;
Matches 247; Conservative 148; Mismatches 345; Indels 118; Gaps 26;
97 DLSIQNREFLFTDCSSKESPPS-IIHOKNGOLS-----LRNKGSMFCNNHAGSGCA 149
113 NLTFSGSLISF-----DSSPTVTYTGQGLSSAGGVNLEIRKLVAGNFTADGCA 187
150 ISADAFSLQHNLYLFTAFENSSKNGGALQAOQFS-LSRNVSPISFARNRADLNGAIC 208
167 IKGASFLITGSGDALFNSNSSSTKGCALATTAGARIANNNGYVPLSNIASTSGAID 247
209 SNLISGVNPLFTTNGSA-TNGAICISDLNTEKGS--LSLACNOETLPASNAKEK 265
248 EGTSLISNNKFLYFEGNAKTGGAIC-----NTKASGSPELIISNNKTLIFASVAVTS 302
266 GGAIYAKHMYLRNGPVSFINNSAKI-----GGATAIOSGSLISLAGGSVLFQNSQRT 321
303 GGAIAHAKKLALSSGGFTFLRNWSSATPKGALSIDASGELISAETGNIIFVANNITLT 362
322 --SDQGLRNALYLEKDALISLSE-ARNGDILFDPPIVOESSKESPLSSLOAVTSPT 378
363 TGSITDTPRRNAINIGSNCKFTELRAKKNHTIFFYDPTISEGSSDYLVKINNGSAGALNPY 422
379 PATASPLVYIOTSANRSVIFSSERLSEEE-KTPDNLTSQLQPIELKSGRLVLDRAVISA 437
423 QGT-----ILFSGETLTADELKADNKLKSSFTQPVSLSGGLLQKCVTLSS 469
438 PSLSDPPALLIMEAGTSKLTSS-DLKLATLSIPLHSLDTEKSVIIRHAPNLSIOKIFLSN 496
470 TSFSGEAGSLIGMOSGTLTSTAGSITITNLGINVDSLGLKQPVSLAKKAS-NKVIYSG 528
497 --SGDENFEYENVELLSKEO-----NNIPLLTSLKESQSHLHPDGNLSS 537
529 KLNLDIDEGNIYES-HMFSHDLFLSKLITVDADVDNVDISL-----IYVPAEDRPS 581
538 HFGYOGDMTFWSKDSDEGHS--LIANWTPKNVYVPEROSTLVANTLWNTYSDMQAVOSM 595
582 EYFGOGQMVNMTTDTATNTEKATATWTGTGFPVSPERKSALVNCNTLWGVFDIRSLQDL 641
596 INTIAHGCAVLEFGTWSAVSNLFYAHDSGKPIDMWHHSLSGLYLGISTHSLDHSFCLA 655
642 VEIGATGMEHKOGFWVSSMTN--FLHKTGDENRKGFRHTSGGYIGGSAHTPKDDLFTFA 699
656 AGOLLGKSSDSFTSTETSYIATV-----QAOLATSLMK----- 690

92 FK-----EKGDLSIONFRLFTDCSSKESPS-----IHK-----NGQLSLRNNGMS 137
109 SSVITONPELCPLSFSGFSOMIFDNCESLSDTSASNVIFHASIYATTPMLFNNOSIL 168
138 FCNRHAECSGASISADAFSLQHNLYLTAPEENSSKNGCAIO-AOTPSLSRNVSPISFAR 196
169 FOYNSAGFGAALRGTSITIENTKSLFNGNCSISNGALGSAIINLINSAPVIFST 228
197 NRADLNGALICS--NLICSGNVNPLFTGNSATNGAICISDLNTSEK-----245
229 NATGIYGAALYLTGSMLTSGNLGSLVFNNSKSGAIYANGVNTSNNSDLTFOONNTA 288
246 -----SLSLACNOETLFAASNAKE 264
289 SPONSLPAPTPPPPPATVPLTGYGALFCTPPATPPPTGVSLLTSGENSVTFLENIASE 348
265 KGAIAVAKHMYLRNGVPSFINNSAKIGALAIQSGSLSLTAGESVLPONNSORTSDQ 324
349 OGGAALYCKKISIDSKSTIFLGNTAGKGAIAIPESGELSLSANQDILFNKNLSITSGT 408
325 GLVRNATYLEKDALLSLEARNG-DILFPDPIVOSSSKESPSPSLQASVTSPTPATAS 383
409 P-TRNSIHFGKDAKFAATGATGCTLYFYDPTSDLSAAS-----AAA 451
384 PLYIOTSANR-----SVIFSEERL-SEEEKTPDNLTSOLOQPIELKSGRLVLDRAVLS 436
452 TVVNPAPKASADGAVSGTIVSGETLTATEATPANATSTLNKLEDEGTALANGATLN 511
437 APSLSDPOALLIMEAGTSLKTS-----LKLATLSIPLHSLDTEKSVTHAPN---486
512 VHNTPODEKSVIYMDAGTTLATNGANNDAITLNLKLVIMLSDIGTKAAVNVQSTNG 571
487 -LSIQIFLNSGDNFENVELLSKEONNIPPLTSLKESGSHLPD-----GNLSHF 539
572 ALTISGTLGLVKNSDCCDNHGMFNKDLQOVPLELKAATSVTTTDDSLCTNGIQOOSFY 631
540 GYGDWTFSMKDSDEGHSILANMTPKNYVPHPEROSTLVANTLMNTYSDMQAVOSMINTI 599
632 GYOGTWEITITTT--HTVTGMKKTKGYLPHPERLAPLIPSLMANVIDLRAVSA--SA 687
600 AHG-----GAYLFGTWGSAVSNLFYA-HDSGKPIDNMHRSLSGLFISTHSLDHSFCL 654
688 ADGEVDPKQSLIT--GITNFFHANHTGDAR--SYRHMGGYLIINTYTRITPPAASL 741
655 AAGOLLKSSDSFISTSTTSYIATVOQLATSLMKIS-----AOACYNESIHELTK 707
742 GFGQLFTKSKDYLVGCHGSHNVFATVYNSITKSLFGSSRFFSGGTSRVYTSKSNKVKTS 801
708 YRSFSEKFGSMHSAVAGEVCAISPIVNSGSL-FSSFSIFSKLQGFSGTDDGEESG 766
802 YTKLPK-GRCSNNSNCMLGELEGNIPILSSRLNLKOLIPVKADEVATHTGGIOENTP 860
767 EIRSFSSAFNISIPLIGITFEKKSQKTRTYVYFLGAYIODIKRDESGPVVLLKANVAS 826
861 EGRIGHGHLNVAVPVGRCGNKSHNRPDYTIIVAYAPDVYRHNPOCDTTLPIGATW 920
827 DAPMANLDSRAVMTLNOALHR-LOTLLNVCYLRCGSHSYSDLGITTYF 878
921 TSGCNLTSTLLVQASSHTSVNDLEIFGHCDCDIRTSROYTLDIGSKLRF 973

RESULT 15
09K2A1 PRELIMINARY: PRT: 995 AA.
AC 09K2A1: 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, last annotation update)
DE POLYMORPHIC MEMBRANE PROTEIN G FAMILY.
GN CP0299.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; Pubmed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,
Berry K., Bass S., Linher K., Weidman J., Khouri H., Claven B.,
Bowman C., Dodson R., Gwin M., Nelson W., DeBoy R., Kolonay J.,
McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL: AE002191; AAF38156.1; -;
DR TIGR: CP0299; -;
SQ SEQUENCE 995 AA: 105599 MW: 8562730FA8B683F CRC64;

Query Match 18.08; Score 816.5; DB 2; Length 995;
Best Local Similarity 26.44; Pred. No. 1,2e-49;
Matches 252; Conservative 158; Mismatches 404; Indels 139; Gaps 25;

37 NPNNVCTFFEDCTMESLFPALCAHAS-----ODDPLVILGNSVCWFVSKLHTDPREAL 91
71 DPKGTLCIFSGDLIYANLDAISRTSSCSFNRAGALQILKGGVFFSLNIRSSADCAAI 130
92 FK-----EKGDLSIONFRLFTDCSSKESPS-----IHK-----NGQLSLRNNGMS 137
131 SSVITONPELCPLSFSGFSOMIFDNCESLSDTSASNVIFHASIYATTPMLFNNOSIL 190
138 FCNRHAECSGASISADAFSLQHNLYLTAPEENSSKNGCAIO-AOTPSLSRNVSPISFAR 196
191 FOYNSAGFGAALRGTSITIENTKSLFNGNCSISNGALGSAIINLINSAPVIFST 250
197 NRADLNGALICS--NLICSGNVNPLFTGNSATNGAICISDLNTSEK-----245
251 NATGIYGAALYLTGSMLTSGNLGSLVFNNSKSGAIYANGVNTSNNSDLTFOONNTA 310
246 -----SLSLACNOETLFAASNAKE 264
311 SPONSLPAPTPPPPPATVPLTGYGALFCTPPATPPPTGVSLLTSGENSVTFLENIASE 370
265 KGAIAVAKHMYLRNGVPSFINNSAKIGALAIQSGSLSLTAGESVLPONNSORTSDQ 324
371 OGGAALYCKKISIDSKSTIFLGNTAGKGAIAIPESGELSLSANQDILFNKNLSITSGT 430
325 GLVRNATYLEKDALLSLEARNG-DILFPDPIVOSSSKESPSPSLQASVTSPTPATAS 383
431 P-TRNSIHFGKDAKFAATGATGCTLYFYDPTSDLSAAS-----AAA 473
384 PLYIOTSANR-----SVIFSEERL-SEEEKTPDNLTSOLOQPIELKSGRLVLDRAVLS 436
474 TVVNPAPKASADGAVSGTIVSGETLTATEATPANATSTLNKLEDEGTALANGATLN 533
437 APSLSDPOALLIMEAGTSLKTS-----LKLATLSIPLHSLDTEKSVTHAPN---486
534 VHNTPODEKSVIYMDAGTTLATNGANNDAITLNLKLVIMLSDIGTKAAVNVQSTNG 593
487 -LSIQIFLNSGDNFENVELLSKEONNIPPLTSLKESGSHLPD-----GNLSHF 539
594 ALTISGTLGLVKNSDCCDNHGMFNKDLQOVPLELKAATSVTTTDDSLCTNGIQOOSFY 653
540 GYGDWTFSMKDSDEGHSILANMTPKNYVPHPEROSTLVANTLMNTYSDMQAVOSMINTI 599
654 GYOGTWEITITTT--HTVTGMKKTKGYLPHPERLAPLIPSLMANVIDLRAVSA--SA 709
600 AHG-----GAYLFGTWGSAVSNLFYA-HDSGKPIDNMHRSLSGLFISTHSLDHSFCL 654
710 ADGEVDPKQSLIT--GITNFFHANHTGDAR--SYRHMGGYLIINTYTRITPPAASL 763
655 AAGOLLKSSDSFISTSTTSYIATVOQLATSLMKIS-----AOACYNESIHELTK 707
764 GFGQLFTKSKDYLVGCHGSHNVFATVYNSITKSLFGSSRFFSGGTSRVYTSKSNKVKTS 823

DR EMBL: U65942; AAC15921.1; -
SQ SEQUENCE 847 AA; 90694 MW; 754C958E7F1179E CRC64;

Query Match 17.2%; Score 779; DB 2; Length 847;
Best Local Similarity 28.3%; Pred. No. 4.6e-47;
Matches 256; Conservative 132; Mismatches 364; Indels 152; Gaps 32;

OY 67 PLYVGNISYCWV--SKLITDPKALFEKGLDLSIONRFLSFTDSSKSSPSITLHOK 124
DB 4 PLY-----WFLISSLIYSN---SLISEPD-----OKTLTAAHY 36
OY 125 NCO-----LSLRN-NGSM-----SFCRNHAECSGAISADAFS-----LOHNYL 162
DB 37 NONTNSEPNPLSTNSNGTITTCGNICIAVAGLDGSLSSCFDITAGNLSFLNGYT 96
OY 163 PFAEENSCKGAGAIQAOFTSLSRVSPIS-----FANRADLNGALCCSMLISGVN 217
DB 97 LCFDNTTOSHPGAIYSGTNKTLDISGFLSCAYCPGATGYAIAVAGTTIKDMS 156
OY 218 NPLFTGNATGGAICICISDLNTSEKSLACNOETLFASNAKEKGAIAKHMVR 277
DB 157 SLVPHKNCSTGEGALQCA---SSSEALKEIENNONLVFAEENSSSGAIAIDKLTIV 213
OY 278 YNGPVSFINNSAKI-----GGAIAIO-SGGSLSILAGEGVLPQNSQRTSDG--LVN 329
DB 214 SGCPFLFSNNVSASPKGAIICIKDSGECSLFDADGDTITFDGNKIKTNGSGPTVTRN 273
OY 330 AYLLEKDALISLEARNG-DILFDPVOESSKESPLSSLOASTSPPAIASPLVIO 388
DB 274 SIDLGSSGKFTKLNAGEGFIYDPTDITGGSD--LNNKODTV----- 316
OY 389 TSANRSVIFSSERLSPEE-KTPDNLTSOLOPTELKSGRLVDRVLAAPSLOPAL 447
DB 317 -DTGKIVFSGERLSEKLVANLKSDFKOPKIKSGSLILDGVTLEKSTOTEGAT 375
OY 448 LIMEACTSLKTSSD---KLATLSIPLHSL-----DTE-KSVTIHAPNLS 488
DB 376 VVMDLGTITQTPSSGETITLNLINVASLGGGVAAPDPAKEATESTVTINAVNL- 434
OY 489 IOKIFLNSGDNFENVELLSKEONNIPLLITSKEOSHLPDGNLS-----SHEYOG 543
DB 435 ----VDNCG--NAVE-YPLIASOFTALEVRSGSGSITKPTTLENYPTTHYCYOG 486
OY 544 DMTFSKDKDEGHSLA--NMPKKNVPHPEROSTLVANTLMTYSQMAVOGMINITIAH 601
DB 487 NMVTWKQSSAOKETATLTWEOTGSPNEROGSLVPTLMGFSDIRAIQNDMSVN 546
OY 602 GCAYLFGTNGSAVSNLFYAHDSGKPIDNMHRSLGLYLFISTHSLDDHFCIAAOLG 661
DB 547 GAYVHGHEFVWSGLGN--FLIKSGSDTKRKRFRHNSAGALGVYAOTPEDEVFAAFCQ 604
OY 662 KSSDSFTTSTETSYATVAOL-----ATSLMKISAQCYNESTIELK 705
DB 605 KKDXYLWKSNSSTVYAGSIYQHIISYWNMTLONTLGAELPLVINAQALAYASNNMK 664
OY 706 TKY-----RSFSGKEGFGMHVAASGEVCASIPISVNSGSLFSSSIFSKLOGFGTO 758
DB 665 TNNDTYAPPKTYYSEKKGWMDGCEVEFGAKAPI-ETISLFLDMKSPYKQLDVIAHQ 723
OY 759 DGEFE--SGGEIRSFASFPNLSLIGITFEKKSQKTRTYVYFGLGAYIOLKRDVESGPV 817
DB 724 DDEKENSDDGRYFESNNLNLNSMPIGVLEKFSHKDTASYNLTLAVALDIVRSNPCTA 783
OY 818 VLKNAVS--WDPAMNLDRAIWEFLTNORALHRLDTLLN--VSCVLRGOSHSLDLGT 874
DB 784 SLVVSPTSAVMTKANLASHAFILQAGNYLALTRNTELFSDQFELRGSCRTYINIDGS 843
OY 875 TYRF 878
DB 844 KIOF 847

RESULT 18
P77792 PRELIMINARY: PRT: 839 AA.
AC P77792;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE POMP90B PRECURSOR.
OS Chlamydomophila psittaci.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
OX NCBI:TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RA Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;
RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RA Longbottom D., Russell M., Jones G.E., Herring A.J.;
RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL: U65942; AAC15921.1; -
KW Signal.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 839 POMP90B.
SQ SEQUENCE 839 AA; 89824 MW; 4581C7CBAE7FFAC4 CRC64;

Query Match 17.1%; Score 776.5; DB 2; Length 839;
Best Local Similarity 28.3%; Pred. No. 6.8e-47;
Matches 246; Conservative 130; Mismatches 349; Indels 143; Gaps 30;

OY 106 LSTFDSSKSSPSITLHOKN-----GOLSLRNGSMFCENHA-----EGSGAISADAFS 156
DB 20 LSFANDQIALTPSDSYNGVNTSEFOVKRTSGTITTCGNVICFAGDSOLKKSCTS 79
OY 157 LOHNYLF-----TAFEE-NSKNGGAI-----QAOFTSLSRVSPISFARNADLNG-C 204
DB 80 ATDNLFTLNGCYTLCFDNITTTASNPICAVOGGKTLGIS--GSLFSCAYCPGTGYG 138
OY 205 AI-CGSMILCSGVNPLFTGNSATNGAICISDLNTSEKSLACNOETLFPASNAK 263
DB 139 AIOTKNTTLKDNSSLVFRHNCSTABGAIQ---KGSSDAELKEIENNONLVFSENSST 194
OY 264 EKGAIYAKHMYRYNGPVSFINN-----SAKIGGAIQ-SGGSLSILAGEGVLPQNS 317
DB 195 SKGAIYADKLTIVSGCFTLFSNNVSNGSSPKGAIISIDSGECLTADLDITFDGN 254
OY 318 SORTSDG--LVNNAIYLEKDALISLEARNG-DILFDPPIVOESSKESPLSSLOAS 373
DB 255 KIITSGSSTVTRNSIDLGTGK-FTKLAKDGFIFDYDITGGGSD-----LNIN 306
OY 374 VTSPTPATPLVYIOTSANSVIFSSERLSEBEKT-PDNLTSOLOPTELKSGRLVYKDR 432
DB 307 KKETVDYTG-----KIVFSGERLSEBEKARENLASFNOPTLSAGSLVLDKG 355
OY 433 AVLSAPLSODPOLLINEAGTSLKTSSD---KLATLSIPLHSL----- 474
DB 356 VSYAKOVTOEAGSTVMDLGTTLQTPSSGGETITLNLININVASLGGGCTSPAKLATN 415
OY 475 DTEKSVTIHAPNLSLOKIFLNSGDNFENVELLSKEONNIPLLITLSK-----EQ 525
DB 416 TASGAILTNVNLV-----DADGNAYED-----PILATSKFTATVATTNA 456
OY 526 SHLPLPGCNLS-----SHFGYQGWTFSMKDSGCHSLIANMTKKNVPHPEROSTLVAN 580
DB 457 STVOTPDNLTNIVPPTHYGGQNMVYWDERTATKATLTWEOTGYSPPNEROGPLVPN 516
OY 581 TLWNTYSQMAVOGMINITIAHGAYLFGTNGSAVSNLFYAHDSGKPIDNMHRSLGLYF 640
DB 517 TLWGAFSDLRAIQLMDISVNGADYHRCGFVWSGLAN--FLIKSGSDTKRKRFRHNSAGIAL 574

RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";
 RL Nat. Genet. 21:385-389(1999).

RC SEQUENCE FROM N.A.

RA MEDLINE-20330349; PubMed-10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RA "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CWL029 from USA."
 RL Nucleic Acids Res. 28:2311-2314(2000).
 DR EMBL: AE001585; AAD18163.1;
 DR EMBL: AP002545; BAA98215.1;
 SO SEQUENCE 922 AA: 100457 MW: DFF2AB633AB031C CRC64;

Query Match 16.3% Score 740; DB 2; Length 922;
 Best Local Similarity 27.2% Pred. No. 3 Le-44;
 Matches 226; Conservative 156; Mismatches 362; Indels 88; Gaps 25;

OY 103 FFFLSFTDCSSKSSPSIIHOKNGQLSRNNGSMSCRNHAECSGGAISADAFLQHNLY 162
 DB 123 FSTLSIQSPGDIKEGCLYSKNA-LMLNNYVRFQONQSKTKGALISGANVTIVGND 181
 OY 163 FTAFEENSSKNGALQAOFTSLRNVSPISEFARNRA-DINGAICC-SNLICGAVNPL 220
 DB 182 SVSFYONATFGCAIHSSGPLQIAVNOAEIRFQONTAKNGSGALVSDGIDIDQNAVYL 241
 OY 221 FFTGNSAT----NGAICISDLTSEKSL-SLACNOETLFASNAKEKGAIVAKHMV 275
 DB 242 FRENELTTAIGKGAVALCCLPTSGSTPVPYVTFSDNKOLVERNHSIMGGAIVARKLS 301
 OY 276 LRYNGPVSEFIN-----NSAKIGGAIAIOSGSLIAGEGVLFQNNNSORTSDOGLVRNA 330
 DB 302 ISSCGTFLINNIVANSQNLGALAIIDTGBEISLSEKCTIFQGN--RTSLPFL--NG 357
 OY 331 IYLEKDAIILSLEARNG-DLFFDPIVOESSKESPLPSSLOASVTSPPATASPLVIOT 389
 DB 358 IHLQNAKFLKQARNGSYIEFYDPTISEAD-----GSTQLINQDPK-----NK 402
 OY 390 SANRSVTFSESERLSEEEKTPDNLTISOLQPIELKSGRLVKDRAVLSAPSLSDOPALLI 449
 DB 403 EYTGTLIFSGEKLAND--PRODKSTIPQVNVNSAGLYVKEGAETVYSKFTSPGSHLV 460
 OY 450 MEAGTSLKTS-SDLKATLISPLHSIDTEKSVTIHAPNLSIOKIFLSNSGD-----ENFY 503
 DB 461 LDGCTLIASKEDIAITGLAIDIDSLSSSTAIVIKANTANKOISVTDSELISPGNAY 520
 OY 504 ENVELLSKEDONNIPLLTSLK-EOSHLLPDCN--LSHFGYOGDWTFSKSDSDE--GHS 557
 DB 521 EDLRM--RNSQTPPLSLLEPGAGSVTVAGDELPSVPHGFGQNMKLAWTGTGNKVGEP 578
 OY 558 LIANMPKNTVPPEPQSTLVANTLWNTYSDMOAVOSMITIAHGAVLYFGTGSASVNL 617
 DB 579 F---WOKINPKPREKEGNLVPAIILMGNAVDSLMQVOETHASSLOTDRGLMIDGIGNF 635
 OY 618 FYAHDSGSKPIDMHHRSGLCYGISTHSLDHSPLAACQLCKSSDSPTSTETTSYI 677
 DB 636 F--HYVASENINIKRNSGIVLSVNNELITPKHTYSMAFSQLFSRDKDVAVSNNERMYL 693
 OY 678 ATVOAQLATSLMKISAQA-----CYNESIHELTKYRS 710
 DB 694 GSYLYOYTTSLGNIIFRASNPNVNVGILSRREFLQNPMLIFHFLCAVGAHTNMKTDYAN 753
 OY 711 FSEYEGCSHMSVAVSGVCSIPVNSGSG-LFSSPSIFSKLOGFSGTQCGFESSGSEIR 769
 DB 754 FPAVK-NSMNNNCWAIECGSMPLVFEENGRLEFGALPEKMLQLVYAVQODEFETADGR 812
 OY 770 SFASASSFRNLSLPIGITEFKSOKTRTYVFLGAVIQLDLKRDVESGPFVLLKNAVSADP 829
 DB 813 RFNGSLITISVPLGIRFEXLALSDOVLVDFSFYIPDIRKDPSCCAALVIGSDSLVLP 872
 OY 830 MANLDSRAVYFRLTNORALHRLQTL---NVSCVLKSGSHSYSLDGTTFYR 878

DB 873 AAHVSHAFVSGTGRYHENDYTELLCRGSIEC--RPHARNINNGSKFRF 922

RESULT 24

ID O9KIY9 PRELIMINARY: PRT: 922 AA.
 AC O9KIY9;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, last annotation update)
 DE POLYMORPHIC MEMBRANE PROTEIN G FAMILY.
 GN CP0770.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AR39;
 RX MEDLINE-20150255; PubMed-10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Uterback T.,
 RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
 RA Bowman C., Dodson R., Gwinn M., Nelson W., Deboy R., Kolonay J.,
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
 RA "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 RT pneumoniae AR39."
 RL Nucleic Acids Res. 28:1997-1406(2000).
 DR EMBL: AE002237; AAF38570.1;
 DR TIGR: CP0770;
 SO SEQUENCE 922 AA: 100480 MW: 3442106888AB0FA6 CRC64;

Query Match 16.3% Score 739; DB 2; Length 922;

Best Local Similarity 27.2% Pred. No. 3 7e-44;
 Matches 226; Conservative 156; Mismatches 362; Indels 88; Gaps 25;

OY 103 FFFLSFTDCSSKSSPSIIHOKNGQLSRNNGSMSCRNHAECSGGAISADAFLQHNLY 162
 DB 123 FSTLSIQSPGDIKEGCLYSKNA-LMLNNYVRFQONQSKTKGALISGANVTIVGND 181
 OY 163 FTAFEENSSKNGALQAOFTSLRNVSPISEFARNRA-DINGAICC-SNLICGAVNPL 220
 DB 182 SVSFYONATFGCAIHSSGPLQIAVNOAEIRFQONTAKNGSGALVSDGIDIDQNAVYL 241
 OY 221 FFTGNSAT----NGAICISDLTSEKSL-SLACNOETLFASNAKEKGAIVAKHMV 275
 DB 242 FRENELTTAIGKGAVALCCLPTSGSTPVPYVTFSDNKOLVERNHSIMGGAIVARKLS 301
 OY 276 LRYNGPVSEFIN-----NSAKIGGAIAIOSGSLIAGEGVLFQNNNSORTSDOGLVRNA 330
 DB 302 ISSCGTFLINNIVANSQNLGALAIIDTGBEISLSEKCTIFQGN--RTSLPFL--NG 357
 OY 331 IYLEKDAIILSLEARNG-DLFFDPIVOESSKESPLPSSLOASVTSPPATASPLVIOT 389
 DB 358 IHLQNAKFLKQARNGSYIEFYDPTISEAD-----GSTQLINQDPK-----NK 402
 OY 390 SANRSVTFSESERLSEEEKTPDNLTISOLQPIELKSGRLVKDRAVLSAPSLSDOPALLI 449
 DB 403 EYTGTLIFSGEKLAND--PRODKSTIPQVNVNSAGLYVKEGAETVYSKFTSPGSHLV 460
 OY 450 MEAGTSLKTS-SDLKATLISPLHSIDTEKSVTIHAPNLSIOKIFLSNSGD-----ENFY 503
 DB 461 LDGCTLIASKEDIAITGLAIDIDSLSSSTAIVIKANTANKOISVTDSELISPGNAY 520
 OY 504 ENVELLSKEDONNIPLLTSLK-EOSHLLPDCN--LSHFGYOGDWTFSKSDSDE--GHS 557
 DB 521 EDLRM--RNSQTPPLSLLEPGAGSVTVAGDELPSVPHGFGQNMKLAWTGTGNKVGEP 578
 OY 558 LIANMPKNTVPPEPQSTLVANTLWNTYSDMOAVOSMITIAHGAVLYFGTGSASVNL 617
 DB 579 F---WOKINPKPREKEGNLVPAIILMGNAVDSLMQVOETHASSLOTDRGLMIDGIGNF 635


```

Db 609 QILISIEITN----DQQLDSNMOSLNTLSLSPHYGQGLMTPNNITTYTTITLNNSS 664
QY 550 -----KDSDEG-----HSLIANMT 563
Db 665 APTSATSIAROKKTSEFTPTSTNTTASIPNIKASAGSGSSASNSGSEVTITKHTLVYNNMA 724
QY 564 PKNYVHPEROSTLVANTLMNTYSDM-QAVOSMT--NTIAHGA--YLEGTWCSAVSNL 617
Db 725 PVGYIDPIRGLDIANSIVSHSGNNMTGRLRLPDNSWFLALQGAATLTFTKQOKRLS-- 782
QY 618 FYAHDSGKPTDKMHHMSLGLGISTHSDHSPCLAAQOLKSSDSDFITSETYSYI 677
Db 783 YHGYSASAK-----GYTVSSQASGAHGKFLSTFSOSSDKMKKETNNLSSRY 832
QY 678 ATY---AOALATSLMKISAOCYNESTHELTKRFSKEGFGSMHSAVSG---EVCA 730
Db 833 LSLALCFHPHEDRIALGAAAC-NYGHNMRSFYGT-KKSKGKFHSTTLGASLRCLRD 890
QY 731 STPLVNSGSLFSSFSIFSKIOGFGTQDGESESGEIRSFSSASFRN-ISLPICITFEK 789
Db 891 SMPLRSLMLPFAQ-ALFSRTEPAS-----IRESGLARLFTLEQAHYAVSPIGIK-- 941
QY 790 KSQKTRITYYFLGAYID---LKRDIESC-----PVLLKKNVSMAPRANL 833
Db 942 -----GAYSSDTWPTLSWEMELAYQPTLYMKRPLNLTLLIONGSMWTTNPL 989
QY 834 DSRAYMFLTNQALHRLQTLNVCVLRGSHSVDLG 873
Db 990 AKHSFYGRGSHSLKFLKLFANQAEVNTVSHYINAG 1029

RESULT 3
YFAL_ECOLI STANDARD: PRT: 1250 AA.
AC P45508: P45507; P45506; P39441; P76468; P77487;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHEICAL 131.2 KDA PROTEIN IN UBIG-NRDA INTERGENIC REGION
DE PRECURSOR.
GN YFAL.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RA MEDLINE=97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA Alpha H., Baba T., Fujita K., Hayashi K., Honjo A., Horiiuchi T.,
RA Ikemoto K., Inada T., Isono K., Isono S., Itoh T., Kanai K., Kasai H.,
RA Kashimoto K., Kim S., Kimura S., Kitagawa M., Kitakawa M., Makino K.,
RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampaio G., Seki Y.,
RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.;
RL Submitted (Jan-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP PRELIMINARY SEQUENCE OF 1-938 FROM N.A.
RX MEDLINE=84272624; PubMed=6087316;
RA Carlson J., Fuchs J.A., Messing J.;
RT "Primary structure of the Escherichia coli ribonucleoside diphosphate
RT reductase operon."
RN [4]

```

```

RP SEQUENCE OF 925-1198 FROM N.A.
RC STRAIN-K12 / EMG2;
RA Estep P., O'Keefe T., Robison K., Church G.M.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1180-1250 FROM N.A.
RC STRAIN-OV6;
RX MEDLINE=88201664; PubMed=2834621;
RA Hussain K., Elliott E.J., Salmond G.P.C.;
RT "The para-mutant of Escherichia coli also carries a gyrA mutation.
RT Mol. Microbiol. 1:259-273(1987).
RN [6]
RP IDENTIFICATION.
RX MEDLINE=96032851; PubMed=7567469;
RA Borodovsky M., McIninch J., Koonin E.V., Rudd K.E., Medigue C.,
RA Danchin A.;
RT "Detection of new genes in a bacterial genome using Markov models for
RT three gene classes."
RL Nucleic Acids Res. 23:3554-3562(1995).
CC -1- SIMILARITY: TO E. COLI YDEK.
CC -1- CAUTION: REF.3 SEQUENCE DIFFERS EXTENSIVELY FROM THAT SHOWN BY
CC MANY FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AEO00313; AAC75293.1; -
DR EMBL: D90855; BAA16052.1; ALT_INIT.
DR EMBL: D90854; BAA16050.1; ALT_INIT.
DR EMBL: K02672; - NOT_ANNOTATED_CDS.
DR EMBL: U30459; AAA74094.1; -
DR EMBL: Y00544; - NOT_ANNOTATED_CDS.
DR Ecogene: EG12850; yfai.
KW Hypothetical protein: Repeat: Signal.
FT SIGNAL 1 23
FT CHAIN 24 1250
FT DOMAIN 919 948
FT CONFICT 28 30
FT CONFICT 40 40
FT CONFICT 65 66
FT CONFICT 431 431
FT CONFICT 433 434
FT CONFICT 478 478
FT CONFICT 773 773
FT CONFICT 853 853
FT CONFICT 923 924
FT CONFICT 948 994
FT PAYOVLNKKYGGYLLNIRANQAEMERRRHHAGGDCQTLN
FT LRVIGG->LTSRC (IN REF. 4).
SQ SEQUENCE 1250 AA: 131152 MW: 177980C5E299FC95 CRC64:

```

```

Query Match 4.4%: Score 197.5; DB 1; Length 1250;
Best Local Similarity 19.7%: Pred. No. 0.0001;
Matches 188; Conservative 143; Mismatches 372; Indels 249; Gaps 38;
QY 53 SLFALCAHASDDPLVVLGNSYCNFVSKLHTIDREKALFKKGLSTONFRLFSFTQCS 112
Db 12 SLFPMINSLFANGVAAVTDSOCGYDVASCOASROSLSGITOWMSINDGQWLVSDFMT 71
QY 113 SKESPSTIHQKNGQLS---RNNGSMSPCRNHAAG---SGCAISADAFLSHNVLTFATF 166
Db 72 NNASCAVFLQGAERSLPENETGTLIFANNVTGGINAGALFRAKENSTL-NLTDVIF 130
QY 167 EENSSKNGCAIOA---OTFSLRNVSPISFARNRA-DLNGCAI-CCSNLSCGNV--- 217
Db 131 SCNVAGYGGALYSSTGNTDGTGAVDLRVTNAMFRNNIANANGKGALEYTINNDVYLSDFVD 190

```

RESULT 5
FIG2_YEAST STANDARD: PRT: 1609 AA.

AC P25653:
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE FACTOR INDUCED GENE 2
GN FIG2 OR YCR089M OR YCR89M OR YCR1102.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;

RN
RP SEQUENCE FROM N.A.
RX MEDLINE=92397594; PubMed=1523899;
RA Wilson C., Grisanti P., Frontali L.;
RT "The complete sequence of a 6146 bp fragment of Saccharomyces
cerevisiae chromosome III contains two new open reading frames."
RL Yeast 8:569-575(1992).
CC
CC - FUNCTION: REQUIRED FOR EFFICIENT MATING.
CC - INDUCTION: BY MATING PHEROMONES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
CC EMBL: X59720; CAA4254.1; -
DR PIR: S19504; S19504.
DR PIR: S25345; S25345.
DR SCD: S0000685; FIG2.
SQ SEQUENCE 1609 AA: 166049 MW: 7066AD7F85A7B852 CRC64;

Query Match 3.78; Score 167; DR 1; Length 1609;
Rest Local Similarity 19.48; Pred. No. 0.013;
Matches 177; Conservative 114; Mismatches 317; Indels 304; Gaps 35;

OY 84 ITDKKELFKKGLDIONFRFLSTDCSSKSSPSTIHKNGQLSLNNGSMFCRHH 143
DB 235 LLSPLLEPLSSSGDLILSTIQATINDOTSK-TIPTLVDAISSLPPLTRSSM----- 286
OY 144 EGSGAISADAFSLQHN-----LTFAPENSSKNGCAIOAQ 181
DB 287 -----APTSGSDSIHNTSPPSKTSNGYDVLTNSIDPSLFTTSEXSS-----TQ 333
OY 182 TFSLSR--NVSPIFARRRADLNGAICCSNLICGNNPLFTFTNSATNGAICISDL 239
DB 334 LSLLRASKSETVNFTASIASTPFGTDSATSL-----DPISSVSTAS--SPVGISFA 385
OY 240 NTSENG-----SLSLACNOETLFASNAKKEGKA-----IYAKH 273
DB 386 NFGTSGNSNYPESTASSSOYODMSSSLPLSQTWVYINTNQGAVTSTTSPAYYST 445
OY 274 MVLRYNGPVS-FI-----NNSAKIGALAIQSGSLIAGEGSVLEFONNSORTSDOG 325
DB 446 ATKYDVGIVTEKVMCPLOTQKSAIGVSSISVPAQSSFG-SLILSSNSTLAASN 504
OY 326 LVRNAIYLEKDAISSLERANDILFDPPIYQESSKESPL-----PSSLAQSVTS 376
DB 505 VP-----ESTASGSSO-----YODMSSSLPLSQTWVYINTNQGAVTS 545
OY 377 PTPATASPLVIOTSNRNVIFSSERLSEEEKTPDNLISLOOPIELKSGRLVLDKRAVLS 436
DB 546 ----TTSRAYVST-ATKY-----DGVITFYVMCPLOTQKSAIGISS-----STIS 588
OY 437 APSLSODPOLLIMEAGTSLKTSIDLKATLISLPLHSIDTERKSVYTIH----- 483

DB 589 ATOTSK-PSSITLIGIST-LQLSDATFKGTETINTHLMTSTSTETETFSGTSDFYLC 646
OY 484 -----APNLSIOKIFLNSGDENFENVLLSKEONNIPLLTSK--EQSHLHPDGNL 535
DB 647 TSEVNLAASSLSVYPNFSSSGSTATITNSVTYFGSTKYSTSVNPTLASQ-----HV 700
OY 536 SSGHGVGDWTFSSKSDSEHSLIAN-----WTPKNVYPHREPOSTLVANTLMNTYSMOA 591
DB 701 SSSVNSLTDF-----SNSTETIAVISNIHKTSNMDY-----SLITTQLKISGKOTL 748
OY 592 VOSMINTIANHGAATLFGTWSAVNSNLFYANDSSCKPLDMNHHRGLVLFGISTSLDHS 651
DB 749 VLSTVTTTVCATVCAATVETTWCPA-SSIAVT-----TSISKTLVLTTEVCSHS 794
OY 652 FCL-----ACQLLCKSSDSFTTST----- 671
DB 795 ECTPTVITSVATSTTIPBLSTSSSTVLSSTVSGAKNPASEVTINTOVSATSEATSTS 854
OY 672 -----ETTSTIATVQAOLAT-----S 687
DB 855 TQVSATSATATASESTTSQVSTASETISTLGTQNTTGTSLPALSTEMINTTVSRK 914
OY 688 LMKISAQACVNESI-----HELKTFRSFSKEGGSMHSAVSGEVC 729
DB 915 TLIISTEVCSHKVCPVITVEVTSKTPSGNSHSSOTLQTEAVEVTLSSHOTVMTSEVC 974
OY 730 AS-----IPVNSGSLGFSFSIFSKLOGSGTODGFEESEGLRFSASFRNLSLP 782
DB 975 SNSICTPVTIVTSVOMRSTFPFVLTSSSTSSSLASTKSSLSLASSEMFTSVST---QSLP 1031
OY 783 IGTFEKKSQKT 794
DB 1032 LAFTCEKERTT 1043

RESULT 6
MSB2_YEAST STANDARD: PRT: 1306 AA.

AC P32334;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MSB2 PROTEIN (MULTICOPY SUPPRESSION OF A BUDDING DEFECT 2).
GN MSB2 OR YGR014W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;

RN
RP SEQUENCE FROM N.A.
RX MEDLINE=92383951; PubMed=1514328;
RA Bender A., Pringle J.R.;
RT "A Ser/Thr-rich multicopy suppressor of a cdc24 bud emergence
defect."
RL Yeast 8:315-323(1992).
CC
CC [2]
CC SEQUENCE FROM N.A.
CC STRAIN=S288C;
CC MEDLINE=97435481; PubMed=9290212;
CC Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
CC "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
chromosome VII."
CC Yeast 13:1077-1090(1997).
CC - I-PTM: O-GLYCOSYLATED IN THE SER/THR-RICH REGIONS (PROBABLE).
CC - SIMILARITY: SOME, TO YEAST HKR1.

CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

```

OY 505 NVELLSKRONNIPULTLSKEQSHLHPDGNLSHFQYOD-----WTFSMKDSDECHSLI 559
DB 357 DL-LVKKSKETYP-----CTEPDISSENDYTCNFNAEESAATSASI 397
OY 560 AMWTKNYPHPEROSTL-----VANTLMNTYSMDQAVOSMINTIAHGAYLFGTQMSAVS 615
DB 398 YSFTPSFENSDISLTSKSKGGEVTFSPYNSGAF--SESNALIMG-----SVS 447
OY 616 NLFYAHDSGKPTIDNMHRSGLYLFQISTHSDHSCFLAAGOLLKSSDFTTSTETTS 675
DB 448 GQRRDTEGS--VNN-----CEINLDNGSTVYVPEVSGKGTNNITSGMLYLH 494
OY 676 YATVQAOLATSLMKISAQACYNESHELKTKRSPSKGEGFSWMSAVSGEYCA----- 730
DB 495 YPDTFTGQ--TVVFKGEVL-----AVDPTETNATPIPVVGYTCKNOIAITADITALSYDG 548
OY 731 --SIPVNSGGLFS-----SFSIFSKLOG--FSGTODGFEESSGEIRSFSSAS 775
DB 549 TTGVLATGCGNRQFPAICGTGFSFSSDFVSSEGFAGAYAYLYNNGVATSAASS 603

RESULT 8
AGAL_YEAST
ID P32323:
AC 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE A-AGGLUTININ ATTACHMENT SUBUNIT PRECURSOR.
RN AGAL OR YNR044W OR N3431.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE-91304412; Pubmed-2072914;
RA Roy A., Lu C.F., Marykwas D.L., Lipke P.N., Kurjan J.;
RT The AGAL product is involved in cell surface attachment of the
RL Saccharomyces cerevisiae cell adhesion glycoprotein a-agglutinin."
RL Mol. Cell. Biol. 11:4196-4206(1991).
RN 12
RP SEQUENCE FROM N.A.
RA Pohl T.M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MEDIATES THE CELL SURFACE ATTACHMENT OF THE A-AGGLUTININ
CC SUBUNIT. S.CEREVISIAE A AND ALPHA CELLS EXPRESS THE COMPLEMENTARY
CC CELL SURFACE GLYCOPROTEINS A-AGGLUTININ AND ALPHA-AGGLUTININ,
CC RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO PROMOTE CELLULAR
CC AGGREGATION DURING MATING.
CC -1- SUBUNIT CONTAINS AT LEAST A BINDING SUBUNIT DISULFIDE-LINKED TO
CC A CORE SUBUNIT.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC AND/OR BY CARBOHYDRATE-MEDIATED COVALENT CROSS-LINKS (POSSIBLE).
CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
CC EMBL: M60590; AAA34382.1;
CC EMBL: 271659; CAA96325.1;
CC PIR: S17031; S17031.
CC PIR: A41258; A41258.
CC SCD: S00053327; AGAL.
CC Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat;
CC Pheromone response.
CC SIGNAL 1 22 POTENTIAL.

```

```

FT CHAIN 23 725
FT DOMAIN 53 493
FT REPEAT 53 149
FT REPEAT 395 493
FT DOMAIN 182 307
FT REPEAT 182 188
FT REPEAT 189 195
FT REPEAT 196 202
FT REPEAT 203 209
FT REPEAT 210 216
FT REPEAT 217 223
FT REPEAT 224 230
FT REPEAT 231 237
FT REPEAT 238 244
FT REPEAT 245 251
FT REPEAT 252 258
FT REPEAT 259 265
FT REPEAT 266 272
FT REPEAT 273 279
FT REPEAT 280 286
FT REPEAT 287 293
FT REPEAT 294 300
FT REPEAT 301 307
FT SEQUENCE 725 AA; 73353 MW; 70420C853B0B01F8 CRC64;

Query Match 3.2%; Score 144; DB 1; Length 725;
Best Local Similarity 19.6%; Pred. No. 0.11;
Matches 102; Conservative 85; Mismatches 223; Indels 110; Gaps 22;

OY 340 SSLRNAGDILFDPPIVQ--ESSKESPLPSLOAVSTPATASPLVIGTSANRSVIF 397
DB 156 SALEPSSASII--SPVISTLSSTSSNPTTSLSTSSPSTSSSTSSSTSSSTSS 213
OY 398 SGERLSEEXT--PDNLTSOLOOPIELKSGRLVLRRAVLSAPLSODPOALLIMAGT 454
DB 214 SSSSTVSSSSSTSSPSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSS 273
OY 455 SLKTS--DLKLATSLPHLSIDTEKSVTHAPNL-----STOKIFLNSGD--ENFY 503
DB 274 STSPSSKSTSAKSTSTSTSTSTSPSLTSSPTLASTSPSTSSSTSTSTSTSS 333
OY 504 ENVELLSKRONNIPULTLSKEQSHLHPDGNLSHFQYODWTFSMKDSDECHSLI 563
DB 334 SSSSTSVLSYSPSTPVYSPVSSNVATPMSST-----VETVSSQSSSE-----YI 381
OY 564 PKNVPRPEROSTLVANTLMNTYSMDQAVOSMINTIAHGAYLFGTQMSAVSNLFY 623
DB 382 TKSSI-----STIPSFMSSTYF-----TVSGVTMYTWTCP-----YSES 419
OY 624 SGRPIDNMHRSGLYLFQISTH--SLDDHSFCLAA--GOLLKS-----SDSFTSTE 672
DB 420 ETSLLTSHNET-----VITDAIVCTHESCHPQOTTSLIISLIMTKTNVATSVSTV 472
OY 673 TTYIATVQAOLATSLMKISAQACYNESHELKTKRSPSKGEGFSWMS 721
DB 473 ESSYACSTCAE--TSHSYSSVQTASSSSVTOQTSTKMSWMSMTTSDDEFKHATGKYH- 529
OY 722 VAVSGEVCASIPVYNSGGLFSSTSPSKLOGFGTODGFEESSGEIRSFSAFRNLSL 781
DB 530 -----VTSSTGSTRIS--TSVEASTSTSIDESQSHLSTLVLSSSSL 575
OY 782 PIGITFEKKSKQKRTYYFFLGAYIODLKRDVESGPVVLK 821
DB 576 TL-----SSDST-----ILFSSVSSL--SVEQSPVTTLQ 603

RESULT 9
YK82_YEAST
ID YK82_YEAST
AC P36170;
DT 01-JUN-1994 (Rel. 29, Created)

```

```

DB 300 SMLDPSQASGSLA-----QGLVPSNEMKOTAIQSPASNTFSHH-----SMVSMIGKC 345
OY 292 GCAIAIOSGGS-----LSLAGEGSLFPNNQORTS--OQGLVRNMIYIE 334
DB 346 FCDVCEAKGSGTTGSOFLQFQTAQALQALAOHS---OSGTTTSSWMG----- 393
OY 335 KQALLSLSEARNGDLFPDPIVOESSKESPLPSLQAVTSPT--PATASPLVITQTSANR 393
DB 394 -----STIOS--PSLVQYOLAKNPSDSAVHSF-----TKR 421
OY 394 SVIFSSERLSE---EERTPDNLTSOL-----QOPIELKSGRLVLDKRAVLSAPSL---SO 442
DB 422 QAFPTSTMEVEFLQEKSPAVATSTAPPPSSPLPSKS-----TSAPOMSGSS 471
OY 443 DQOALLIMEAGSLKTSDDLKATLSLPLSHLDTEKSVTHAPNLISOKIFLSNGDENF 502
DB 472 DMOSSSPOPAQOKLKQOKKASLTSTKIPALAVMPCGADISGLNLOPGALQFCSEPVLSO 531
OY 503 YENVLLSKQNNIP--LLTLSEKQSHLHPDGNLSHFQYQ-----DWTFSWKDSDEC- 555
DB 532 YESTPTTSSSSQAPSSLTSTASSESSSTI--SSNOQESQYOGPIOSTTYTSQNNAGCP 590
OY 556 ----HSLIANMTPKNVYPPPEROSTLVANTLMNTYSDMOAVOSMINTIAHGAYLFTGWC 611
DB 591 LVEQRSTOTRRPYSISISSFOKDLTOAK-----NGFSSVQATQLOTTQSVEGAT-----G 640
OY 612 SAVSNLFYAHDSGKPID-----NMHRSIGYLFQISTISLD-----DHSF 652
DB 641 SAAVKS--DSPTSTSIPLNETVSASALLTTTNOHSSSLGGL---SHSEEPNPTTTTQHS 695
OY 653 CLAAGC--LLGKSSDSFTSTETTSYIATVOALATSLMKISAQACNESIHELKTYRSF 711
DB 696 TLTQONTLSSSTISGRTISTL-----LHTSV-----ESANLHSSSTF 736
OY 712 SKGFGSMHSAVSGEVCASIPVNSGGLFSSFSIFSKLQGFSGTODGEESGGEIRSF 771
DB 737 S-----TTSTSVASPPVSVSSSLNSGSLGLSL--GSNSTVTASTRSSVATTSG 785
OY 772 SASSFRNISLPTGI 785
DB 786 KAPP-----NLPGV 795

RESULT 11
AMT_CLOAB STANDARD: PRT: 469 AA.
ID AMY_CLOAB
AC P23671;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE PUTATIVE ALPHA-AMYLASE (EC 3.2.1.1) (1,4-ALPHA-D-GLUCAN
DE GLUCANOHYDROLASE) (FRAGMENT).
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 792;
RX MEDLINE=91072241; Pubmed=2254264;
RA Gerischer U., Duerre P.;
RT "Cloning, sequencing, and molecular analysis of the acetacetate
RT decarboxylase gene region from Clostridium acetobutylicum.";
RL J. Bacteriol. 172:6907-6918(1990).
CC -1- CATALYTIC ACTIVITY: ENDOMYDOLYSIS OF 1,4-ALPHA-GLUCOSIDIC
CC LINKAGES IN OLIGOSACCHARIDES AND POLYSACCHARIDES.
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M55392; AAA63759.1; -
DR PIR: B37837; B37837.
KW Hydroxylase; Glycosidase; Carbohydrate metabolism.
FT NON-TOXER
SQ SEQUENCE 469 AA; 50870 MW; DB1B6702D9B01572 CRC64;

Query Match 3.0%; Score 138; DB 1; Length 469;
Best Local Similarity 19.0%; Pred. No. 0.14;
Matches 142; Conservative 81; Mismatches 214; Indels 312; Gaps 36;

OY 124 KNGQLSRNNGSMSCFRNNAEGSGAISADAFSLQHNYLFTAFEESSKNGGAIQATF 183
DB 4 KSGDLT-----NAQMGGLDSKKCYDMLETH--DEYHNESK----- 38
OY 184 SLSRNVSPISFARNRADLN-----GGAICCSNLICSGVNPFLF---TGNATNGAIC 234
DB 39 -----DLTDQKAGMALIASR---AGSV--PLFFDRPTGNIGSEGDALM 78
OY 235 CISDLNTSEKGSLSLACNOETTFASNSAKKGAIIAKHVLVLRNGPVSTINSAKIGA 294
DB 79 KDSDVAVVNEFHNAAGONEYLRLO-----NNKAMI----- 110
OY 295 IAIOSGGSILAGEGSLVLENNQNSORTSDQGLVRNMIYIEKDAIISLEARNGLDIFDP 354
DB 111 --IERGSGAVIVNEGSEFLNTP-----INLEGND----- 141
OY 355 IVOESSKESPLPSLQASVTSPTPATASPLVITQTSANRSVIFSSERLSEEXTPDNLTS 414
DB 142 --NHGATDLSLVS--QGRNTGTVPANSIIVLYNKNNSNP----- 177
OY 415 QLOOPIELKSGRLVLDRAVLSAPLSODQALLIMEAG--TSKLTSSDLKATLSLPLHS 473
DB 178 -----SDRVTLSQOAKKAGDSVT-----ITYDAGTALKDASNVNL-----YWG 216
OY 474 LDTEKSVTHAPNLISIOKIFLSNGDENFENVELLSKEQNNIPPLTLSEKQSHLHPDG 533
DB 217 YDGFSAATSKA-----MTSLGDKKQOTTIVPEVTKNV-----NFSSTDG 257
OY 534 NLSHFGYQGDWTFSMKDSDEGHSILANWT---PKNVYPH-----PEROSTLVANTLMNT 585
DB 258 T-----SW--DNNGG-----ANWNIPLASNYLPHGGYKVYDDSSNLVSGNNFTI 299
OY 586 YSDMOAVOSMINTIAHGCAVLFGTWC-----SAVSNLFYAHDSGKPIDNMHHSLSGLFG 641
DB 300 YVNGNLANSS--NVSLLH-----WGVNGWSNMOLAMVYDSNC-----FW 336
OY 642 ISTHSLDDHS-----FCLAAGLLKSSDSFTSTETTSYIATVOALATSLMKISAQAC 697
DB 337 EATTAIPASSMTLNFCTTNGSSMDNNNNNNMT--LNTSSVYPKVQ----- 379
OY 698 NESIHELKTYRSFSGEFGSMHSAVSGEVCASIPVNSGS--GLFSFSIFSKLQGF 755
DB 380 -----VTAPAECQKISVYVNGSLASASNITLHMGCGFT 415
OY 756 GTODGEESGGEIRSFASASSFRNISLPIGITEFKSKQKTRTYVFLGAYIIDLKRDYEG 815
DB 416 SPQD-----INMVKQADGRMLANITLPSGC-----YVNVMAFMD-----QSG 452
OY 816 PVLILKANVSADAMANDSRAVYMERLTN 844
DB 453 -----TWD-----NNNSNNVNSSSTN 468

RESULT 12
HKRL_YEAST STANDARD: PRT: 1802 AA.
ID HKRL_YEAST
AC P41809;

```

RT Characterization of the gene encoding the protective paracrystalline-
 RT surface-layer protein of Rickettsia prowazekii: presence of a
 RT truncated identical homolog in Rickettsia typhi.
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-BREINL;
 RA Moron C.G., Yu X.J., Walker D.H.;
 RT "Sequence analysis of ompB of Rickettsia prowazekii."
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN-MADRID E;
 RX MEDLINE-99039499; PubMed-9823893;
 RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,
 RA Sichteritz-Ponten T., Almarik U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria."
 RL Nature 396:133-140(1998).
 RN (4)
 RP PARTIAL SEQUENCE.
 RC STRAIN-BREINL;
 RX MEDLINE-92114896; PubMed-1370573;
 RA Ching W.M., Carl M., Dasch G.A.;
 RT "Mapping of monoclonal antibody binding sites on CNBr fragments of
 RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
 RT prowazekii."
 RL Mol. Immunol. 29:95-105(1992).
 RN (5)
 RP IDENTIFICATION OF CLEAVAGE SITE.
 RX MEDLINE-92104668; PubMed-1729180;
 RA Hackstadt T., Messer R., Cieplak W. Jr., Peacock M.G.;
 RT Evidence for proteolytic cleavage of the 120-kilodalton outer
 RT membrane protein of rickettsiae: Identification of an avirulent mutant
 RT deficient in processing."
 RL Infect. Immun. 60:159-165(1992).
 RN (6)
 RP FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC - FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
 CC - CELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
 CC LAYER WITH HEXAGONAL SYMMETRY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M37647; AAA26390.1; ALT_INIT.
 DR EMBL: AF161079; AAD42234.1;
 DR EMBL: AJ235273; CAA15140.1;
 KW Antigen; S-layer; Cell wall.
 FT CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1 1329 1643 32 KDA BETA PEPTIDE.
 FT VARIANT 257 257 V -> A (IN STRAIN BREINL).
 FT VARIANT 1010 1010 Y -> D (IN STRAIN BREINL).
 FT VARIANT 1450 1450 A -> S (IN STRAIN BREINL).
 FT VARIANT 178 179 AA -> VC (IN REF. 1).
 FT CONFLICT 191 201 TTOTAPLTLGA -> INSRSSHYLVS (IN REF. 1).
 FT CONFLICT 212 212 T -> I (IN REF. 1).
 FT CONFLICT 313 313 Q -> L (IN REF. 1).
 FT CONFLICT 1104 1104 D -> G (IN REF. 2).
 FT CONFLICT 1123 1123 T -> S (IN REF. 2).
 SEQUENCE 1643 AA; 169854 MW; 735FDF392E6346CC CRC64;

QY 130 LNNNGS-----MSECRNHAEGSGAISAADAFSLQHNLYLFTAFEESSKNGCAIOAOTFSL 185
 Db 424 VKNNNGTAVGITFENAN-----GALVAST-----DENIVATNIAEAGAGV 466
 QY 186 SRNVSPISFARNRADLNGAI---CCSNLCSGVNPLFFTGNGATNGAICCSIDNTSE 243
 Db 467 V-ELSGIHAEIRLG-NGGSIFFKLADGTIVNGPVNONALNNNAAGSI-----QLDGSA 520
 QY 244 -----KGSLSACQNETLFASSAKKEGCAIYAKHMYRNGPVSFINNSAKIGGAIA 296
 Db 521 IITGIDGNGVNAALQHTL-ANDASK-----ILALDG-ANLT--GANVGAIH 565
 QY 297 IO-SGGSLSILAGEGYLFPONNSORTSDGLVRNATYLEKDAILSLSEARNGOILEFPDPT 355
 Db 566 FQANGGTIKLNTQNNINLVNFEDLDTDKTGVDASSLTNNQTLTI-----NGSI----- 615
 QY 356 VOESSKESPLPSSLOASVSPPTATAPLYIOTSANRSVIFSSERLSEKTPDNLTSQ 415
 Db 616 -----GTVAANTKTTLAQ-----LNIGSSKTLNADVAINELVIEENGSVQ 656
 QY 416 LQPIEL--KSGRLVKDRAVLSAPSLSDPQALLMEAGTSLKTS-----SDLKATLSI 469
 Db 657 LNHNTYLTITKT INANOGQILVADPLNTN-----TLADGTNLGSAENPLSTIHFAKAA 712
 QY 470 PLHS-LDTEKSVTHAVNLSTOKIFLSNGDENFENVVELSKBONNIPLLTJSKEQSH- 527
 Db 713 NADSLVNGVGNLYANNITND-----ANVGLHRSRG-----GTSIVSGTVGGQGHK 762
 QY 528 ---LHLPDGNLSHFYOGDWTFSKDSDECHSL--IANTPKRYVHPPEQSTL-VANT 581
 Db 763 LNNLILNGTIVK---FLGDTTFNGGTRITECKSLIOISNNYTTDVESADNTGTLFVNT 819
 QY 582 LWNYSQMAVQSMINTIAHGAGLFGWGSVAV-----SNLFYAHDSGKPIDNMHHRSLSG 637
 Db 820 -----DPTVTILNKGAY-FGLYKQVYIISGQNIYF-----NEIGN-----VG 856
 QY 638 YLFGTSHSLDHSFCIAACQALGKSSDSPTSTETTSYATVQAOALATSLMKISAQCY 697
 Db 857 IVHAIANSISFENASLGLSLFLPSGTPLDLITKSTGVNTQVNFNAPIYVVGSDSMI 916
 QY 698 NESIHELKTKRSTSKBGFSGWHSVAASGEVCASPIYNSGSLFSSSISIKSQPSGT 757
 Db 917 NNG--QLIGDKKNIILALSGDSNITV-----NANTLYSGIRTKNNGQIVTL 962
 QY 758 QDGEESGCELRSSASSFRNISLPITFEKKSOKR-----TYYPFLAYIO----- 806
 Db 963 SGMKPNNGFTI-----YGLGLENGSEKFKOVFTTDYNNLGSIIANNVTIN 1008
 QY 807 -----DLK-----RDVES-----GPVLLKNA 823
 Db 1009 DVTITTGICGIDFDAITILGSVNGNANVFVDSTFSDPSMIAVATQANKGYTYLLNA 1068
 QY 824 V-----SMDAPMANLDSRAVFRLTNQRALHRLQTLNLVSCYLKQSSYSYSLDGT 874
 Db 1069 LVSNIGSLDTPVASV-----RFTGN-----DSGAGLOGNIVSONIDFOT 1107

RESULT 14
 FAB-MOUSE
 ID 006194; STANDARD: PRT: 2319 AA.
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE COAGULATION FACTOR VIII PRECURSOR (PROCOAGULANT COMPONENT).
 GN CF8 OR F8C.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN (1)
 RP SEQUENCE FROM N.A.

Query Match 3.0%, Score 134.5; DB 1; Length 1643;
 Best Local Similarity 19.9%; Pred. No. 1.6;
 Matches 166; Conservative 128; Mismatches 299; Indels 243; Gaps 41;

CC - SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCF/TPX1;
 CC INSDCTS AG3/AG5; FUNC1 SC7/SC14 AND PLANTS.PR.1
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z49353; CAA89370.1; -
 CC EMBL: X83502; CAA58492.1; -
 CC EMBL: X88851; CAA61314.1; -
 CC HSSP: P04284; ICPE.
 CC SCD: S0003614; YJL078C.
 CC InterPro: IPR001283; -
 CC Pfam: PF00188; SCP.1.
 CC PRINTS: PR00837; V5TPYLKE.
 CC PROSITE: PS01009; SCP_AG5_PRL_SC7_1; 1.
 CC PROSITE: PS01010; SCP_AG5_PRL_SC7_2; 1.
 CC Hypothetical protein.
 CC KW DOMAIN 166 379 ALA/SER/THR-RICH.
 CC SEQUENCE 881 AA; 89152 MW; 37DBAC660CA9D12A CRC64;
 SO
 Query Match 2.8%; Score 126.5; DB 1; Length 881;
 Best Local Similarity 17.9%; Pred. No. 2;
 Matches 152; Conservative 147; Mismatches 328; Indels 223; Gaps 37;
 Oy 95 KGDLSIQNFRRLSTPDCSKESPSTIIHQKNGOLSLRNNGSMSECRNAEG--SGA1S 151
 Db 17 KAOTTFPME---SDVLNHNKRALVNDTAPLTM-SDTLATYQONADYDCGVLTH 71
 Oy 152 ADAFSLQHNLYLFTAFENSSKQ--NGAIOAQTSLR-NVSPISFA-----RN 197
 Db 72 SDG-----PYGENLALGYDTGTAVDAVMGEISKYVNSNPGFSESTGHFTQVWKS 121
 Oy 198 RADLNGAICCSN-----LJCSGN-----VNPFLTGSATNGAIC-----C 235
 Db 122 TAEICGCKYCGCTWNNYIVCSYNPGNYLGEFAEEVEPLISTVSSSSSSSTSTSDT 181
 Oy 236 ISDLTSEKGSILACNOETLFASSAKEGAIYAKHVLVYNGPVSPINNSAKIGAI 295
 Db 182 VSTISSIMPAVAGYTTVSSAASSSLKSTTINPAKATLTASSSYITTSYVGS 241
 Oy 296 AIOGSGSLI--LAGEGSLVFQNN--SQRTSDQGLVNAIYLEKDAIISLEARNGITL 350
 Db 242 TVSSASSSVTTSYATSSSTVSSDATSTTTSSVATSSSTSSDPTSTAASSDPA 301
 Oy 351 FPDPIVQESSKESPLPS--LQASVTSPTPATAPLVLTQTSANRY----- 395
 Db 302 SSSAASSASTENAASSSAISSSVAPLSSTLTSTASSSRVSVSVSVKFAANT 361
 Oy 396 -IFSSRLSE-----EETPDNLTSLOQDPIELKSG-----RLVLDK 431
 Db 362 TVFSADTTSSVASLSSVAADDIQGTISKESVSEHTSYTATNAQAOTRIGSSS 421
 Oy 432 RAVLSAPSLQDPOALLIMEAGTSLKTSDLKATLSLPLHSL-DTEKSVTHAPNLSIQ 490
 Db 422 RSSSGAVSSSVASQVSL-----NSVIAVNTDVSTSVSTAHTTKDPATTSVASESITSE 477
 Oy 491 KIFLNSGOENTFENVELLSEKQNNIPILLTLEKQSHLHLPDGNLSH-FGYGQDMTFSH 549
 Db 478 TAOASSSTENKINSATSSITSYSN-----SASVSGHGVTAAYEAIRT- 520
 Oy 550 KQSDDEGHSILANTPKNYVHPEROSTLVANTLMN--TYSOMAQVQSMINTIANGAVLFG 608
 Db 521 --SEOSSALATS-----VPATNCSSIVKTTTLENSSTTTITATITSTTLA----- 564
 Oy 609 TWGSAVSNLFYA-----HDSGKPIQNMHRSIGYLFQIST--HSIDD-----H 650
 Db 565 ----TANSTRATATVITDPTLPDONSASPTONAKHTS--TYGSSSTGASLDSLTFTT 618

Oy 651 SFLCAG--OLG-----KSSDFITSTETTSYATVQAOALATSLMKISAQACYNES 700
 Db 619 SIVSSNTQVLVECTGCTSESDYSDSPFAISTATTT-----ESMLINTLNITAS--CSTDS 670
 Oy 701 IHEIKTVRSFSEKSGRSHSVASGEV--CASI-----PIVNSGSLFSS 744
 Db 671 -----NFTPSAASSTDETAFTTRISTSCSLNGASTQTSLETTSPKMTVTPASS 721
 Oy 745 F-----SIFSKLOGFSGTODGFEESSGEIRSPASAFRMSIPGITFE 788
 Db 722 FPSITTTCLENDTAFSSITTEVNAATINPG--ETISLASDRATSKREPRETSVASTSN 779
 Oy 789 KKSQKTRTY 798
 Db 780 EGTSSITTTY 789
 RESULT 23
 YFL9_SCHPO STANDARD; PRT: 1379 AA.
 ID YFL9_SCHPO
 AC 014066: 09UOM4:
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL 150.0 KDA SERINE-RICH PROTEIN C1687.09 IN CHROMOSOME 1.
 GN SPAC1687.09 OR SPAC2211.01C.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 NX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Wood V.;
 RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AL035064; CAA22603.1; -
 CC EMBL: AL031181; CAA20148.1; -
 CC Hypothetical protein.
 KW DOMAIN 1351 1358
 FT SEQUENCE 1379 AA; 150016 MW; 9720ADEA099C56E8 CRC64;
 SO
 Query Match 2.8%; Score 126; DB 1; Length 1379;
 Best Local Similarity 18.6%; Pred. No. 4.2; Indels 326; Gaps 40;
 Matches 171; Conservative 120; Mismatches 301;
 Oy 151 SADAFSLQHNLYLFTAFENSSKQ--NGAIOAQTSLSRNVSPISFARNRADL----- 201
 Db 232 AAFQSLREDFDTTPPEPSDKLEKLNKASQASQESQSLKMEISLARSPLITSEK 291
 Oy 202 --NGA-----ICSNLIGSGVNPPLFTGNSA--TNGAICISDLNTSEKSL 247
 Db 292 LKNKASKQALIESPPFASERPLSSNIIP-----NPMERLKNASKLATIESQPFKSAEPL 346
 Oy 248 SLACNOETLFASSAKEGAIYAKHVLRYNCPV-----SF--INNSAK-- 290
 Db 347 SSAT--PLPNPSEKMRNCA--SKOAIMAOSKINPLPLPILASISDPSEKIKLNTASKO 401

Query Match 2.8% Score 126; DB 1: Length 1589;
 Best Local Similarity 18.1%; Pred. No. 5.2;
 Matches 162; Conservative 141; Mismatches 322; Indels 272; Gaps 37;

112 SSKSSPSPIHOKNOLSLRNGSMFCNNAHEGSGAISAQA-PSLONNYLFTFEENS 170
 143 SSKSSRRSLN-----SLGNSAYLHVPRNPSKSRGSSSTLASLSMAIN-----AETSS 191
 171 SGNCGAIAQAFSLSRNVPISFARNRADLNGAICCNLICSGVNPFLFTGNSATNG 230
 192 GHNNTYVSNMNSPFSAPNDASHIT--PSSNFMNSMNSLSDMTKRSAD-----GSEENMT 242
 231 GAICCSIDLNTSEKSGSLACNOETLFASNSAKKGAIAAHMV-----LRYNCP 281
 243 NA1-----MNNNE-----TNLQT-----SGEKAGPPIVAETIKILPLEETEMIING- 284
 282 VSFNNNAKIGCAIAIQSGSLSLIAGEGVLFONNSORTSDOGLVRNMIYLEKD----- 336
 285 --IRSNIASTWSPILPLIT-----RTSDRYLV-----YNNKDLIYC 318
 337 ---ALISLEARNQDILFTDP-----IVQESSKESPLSSLOASVTSPTPTAS 383
 319 SELPLISNIMESDDICDEPKPPPDHLVNLTRDLRKNANIEDSSTSKOSEQONS 378
 384 PLVIQTSANRSVIFSSERLSEEEKTPDNLTSOLOQ--PIELKSGRLVLDRAVLASPSLS 441
 379 SULMEKQDKETDGNNGNSINDONNNNNKNEFENAGPSSLNS-----LSAPDLT 428
 442 QDPQALLMEAGTSLKTS-----DLKATLSTPLSLDTEKSVTHAPLWLSOKIF 493
 429 QNIOQSVVAPSRSSIIAKSDIFHYNSRIDIKLMT--ELDOLTYVYKTAH-----KMF 478
 494 LMSNGENYEVENELLSKE-----QNNIPLTLTSKE-----Q 525
 479 LKEN-RUNLFTKYRDLISDSIVFTOLGCRMLQHEIKAKSCKEIKKIFKGLISSLSRISIN 537
 526 SHLPLDGNLSHFYQGDWTFMSKDS-----EG-----HSLIA----- 560
 538 SHLY-----FDSA FHRKKMDTMNDKNDQENNCSPTEGDDCKIEVDVHDLVSPLSGK 592
 561 -----NMV-----PKNVVPRKOSTIVANTLNTYS--- 587
 593 RNVTSTDTLTPMRSSFTVNENDEMNFVLPBNSVNSVWTPRISTONSTLEDSPSN 652
 588 ----DMQAVOSMINTIAHGCAVILFTGSAVSNLEVAHDSGKPRIDMHHRSGLYLE--G 641
 653 KNFKSAKSTIEMDV-----EBSKFLRHVOLLIFVLQSSVSDDNTLQLLPREFKGS 705
 642 ISTHSLDHSFLCAAGOLLKSSDSFTTETJ-----SYIATVQALATSLMKISAQA 695
 706 FSGSGWNPSTFTTDFGNATKKKAVTSNBDVTAASSKNSISIRPKMADALIASAGYS 765
 696 CYNESHEHLTK-----YNSFEKGFSGMSHVAVSGEVCASITVNSGSLF--SS 744
 766 ANSETNSQIDLKASSAASGVFTPFNRPSSHRTFSRAVSRKRRKRYLVTDLTMTMKKS 825
 745 FSTPSKLOGFGSQDDFEESGGEIRSPASASFRNISLPDITGFEKKSKOKRTYVFLGAY 804
 826 SOTFEKLNATG-----EHLKTIKPKSRIRHLEIN-SSTYQINO-----NVL 868
 805 IODLRDVEGSPVLLKNAVSWDAPMANLDSRAYMFRLLNORALHRLQTLNNSCVL 861
 869 LLELLENLDSLIFNLKLNLIKTPSILLDLSEEFLL-----VHAMSVSVSVL 914

DE HOMEOBOX PROTEIN SIX4 (SINE OCULIS HOMEOBOX HOMOLOG 4) (SKELETAL
 DE MUSCLE-SPECIFIC ARE BINDING PROTEIN AREC3).
 GN SIX4 OR AREC3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C; TISSUE=Skeletal muscle, and Myoblasts;
 RX MEDLINE=96178400; PubMed=8628654;
 RA Kawakami K., Ohto H., Ikeda K., Roeder R.G.;
 RT "Structure, function and expression of a murine homeobox protein
 AREC3, a homologue of Drosophila sine oculis gene product, and
 RT implication in development.";
 RL Nucleic Acids Res. 24:303-310(1996).
 CC -I- FUNCTION: INVOLVED IN SKELETAL DEVELOPMENT. ALSO IMPLICATED
 CC -I- IN RETINA AND KIDNEY DEVELOPMENT.
 CC -I- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.
 CC -I- ALTERNATIVE PRODUCTS: AT LEAST THREE ISOFORMS OF THIS PROTEIN ARE
 CC PRODUCED BY ALTERNATIVE SPLICING. M8 CONTAINS NO HOMEOBOX DOMAIN
 CC AND IS NOT FOUND IN MUSCLE.
 CC -I- TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING SKELETAL MUSCLE. ALSO
 CC PRESENT IN ADULT RETINA AND IN THE DISTAL TUBE OF KIDNEY.
 CC -I- SIMILARITY: BELONGS TO THE SIX/SINE OCULIS FAMILY OF HOMEOBOX
 CC PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D50416; BAA08915.1; -
 DR EMBL: D50417; BAA08916.1; -
 DR EMBL: D50418; BAA08917.1; -
 DR MCD: MGI:106034; SIX4.
 DR InterPro: IPR001356; -
 DR Pfam: PF00046; homeobox; 1.
 DR PROSITE: PS00027; HOMEOBOX_1; 1.
 DR PROSITE: PS0071; HOMEOBOX_2; 1.
 KW Developmental protein; Homeobox; DNA-binding; Nuclear protein;
 KW Alternative splicing.
 FT DOMAIN 2 5
 FT DOMAIN 58 61 POLY-SER.
 FT DOMAIN 70 76 POLY-ALA.
 FT DOMAIN 92 95 POLY-ALA.
 FT DNA-BIND 216 275 HOMEOBOX.
 FT DOMAIN 582 775
 FT VARSPLIC <1 9
 FT VARSPLIC 37 100
 FT VARSPLIC 188 319
 FT VARSPLIC 320 775
 FT SEQUENCE 775 AA; 82262 MW; B06EBB64E04E5061 CRC64;
 SO

Query Match 2.8% Score 125.5; DB 1: Length 775;
 Best Local Similarity 20.5%; Pred. No. 1.9;
 Matches 147; Conservative 67; Mismatches 219; Indels 283; Gaps 37;
 3 PDHNFCCIAAI-----LSSTAVLFGQDPLGETALLTKNHNVCCTFE 47
 103 PDHV-ACVCEALQGGNLDRLARFLWLSLPQSDLLRGNSILLKARALVA-----FH 151

```
OY 48 DCTMESLFALCAHA--SODDPLVYLGNSYCEVSKLHTDPKEA----- 90
Db 152 OGIVPELYSILSHSFESANAPLL---QOLWY--KARYTEARARPLGAVDKYRLRR 205
OY 91 -----LFEKEG-----DLSTIONFRF-----LSFTDCSS-- 113
Db 206 KFLPLRTIMDEEETVYCFKEKSRNALKELYKON-RYPSPAEKRLAKITGSLTOVSMMF 264
OY 114 -----KESSPSIIHQKNGQLSRNNGSMSFCRNHAGSGGAISADAFSLQHNLYLFTAFEE 168
Db 265 KNRQRODRNPSETQSK-----SESDGNPSTEDESSKGH-----ED 299
OY 169 NSSKNGGCAIOAOT-FSLSRNVSP1---SFARNRADLNGAICCSNLICSGNVNPLEFTTG 224
Db 300 LSPHPLSGASDGVNLSLSHVEPYMOQIGNAKISLSSGVLLNGSLVPASTSPVFLNG 359
OY 225 NSATNGCAICICISDLN-----TSEKGSILACNOETLFAS 259
Db 360 NSFIOGHNGVILNGLNGNTQTVSLNPPKMSNIVNGIAMTDIGSTSDVKEKVLQS 419
OY 260 NSAKEKGCAIYAKHMYLVRYNGVPSPFN-----NSAKICG--AIAIOSGSLSLAGEGCVL 313
Db 420 SAVNSAATTSP-----SAPVSPGLIPCTEVKREGIOIVASQDGGSV-----VT 465
OY 314 FQNNQSQRTSDQGLVR-----NAIYLEKDAILSLLEARNGDILF---FDP1-VOESSSK 362
Db 466 FTTPVO-INOYGIQVOPNSGANGQFL-----NGSIGFSPLOLPPVSVASOGN 512
OY 363 ESPLSSIAQAVTSPTPATA-----SPL-----VIQTSANRSVIF 397
Db 513 LSVTPSTSDGSTFTSEPTVHGKLFSLTPSAVVYTVPNSGOTVGAVKQDEGLERGLVF 572
OY 398 -----SSERLSEEEKTPDNLTSQ-----LQOPTELKSG 425
Db 573 SOLMPVNHSAQVNAHSLSENLSGGLHP--LTSSLVNVSAHGFSLTPPTLLNPTELMPD 630
OY 426 RLVLKDRAVLAP-----SLSDPPQALLIMEAGTSLKTSDDLKATLSIPL----- 471
Db 631 ---LAESQPVAPVASKCTVSSVSNTNYATL--ONCSLIPGODL---LSCPMTOAALG 680
OY 472 -----HSLDTEKSVTIHAPNLSIOKIFLSNSGD--ENFYEVELLSKEONNIPLL 519
Db 681 EIVPTAEQVSHASTAVHQDFVREORLVLOSVPNIKENFLONSF--NKATNNLMML 734
```

Search completed: May 6, 2001, 19:28:01
Job time: 165 sec

```

OY 291 -IGGAIATOSGSL-----SILAGESSVLPFONNSORTSDOGLVRNAI-----Y 332
DB 402 ALESOSLNNKSGELLEPSISDPMLEKMTDACKQTESLSLKSDDATIPKPSVSLGVK 461
OY 333 LEKDAIILSLERN-----GDILFPDPIVOE----- 358
DB 462 LKNDASKCAVESQSLMINDPSVSTSFYPMNEKLNKAIOAIEHOALNAAILNKTOLP 521
OY 359 ---SSKSPSPSSLOASVTSPTATA----- 382
DB 522 YFHQSSSELPISAKRAALQOTESASKSSNISSEKCDHPSPNSFISASQOAKDRFESE 581
OY 383 ---SPLVIOTSNARSVIFSSERISEE-----KPPDN-----LTSOL 416
DB 562 ATSSIDLDELSPGRSLASKLSTVKEAGARYNRSVAPVPMPSPNPFELSKASQSLASTA 641
OY 417 COPIE--LKSGLVLKDRVLASPSISODPOLIMEAGTSLKTS-----SDLK 463
DB 642 OVPVEPAPKTSSEVNLKSLSLASKLS-----FOEAGTSLKERTANYQHSPEVLNDRS 694
OY 464 LATTISPLH---SLDTEKSVTHAP-----NLSTOKIFLSNGDENEFYE- 504
DB 695 LASTLSARHADAISSSKOHTEIADPAFMTGTGLMTKTKNSLSDSALTPQHSDSVLPDI 754
OY 505 -----NVELSKRONNIPILITISKOSH-----LHLPDGNLSH- 538
DB 755 TSKAARLSTCAKTDLHRPRKHKSTLENYFGHNAEDESPQSDSEVESQTPPGSYSEND 814
OY 539 -----FGYOGDWTFSWKD-----SDECHSLIANWTP-KNVVPREROSTLVANT 581
DB 815 ICGTDIYFPALOGAALAAKDFSRREASISDSQASASMDLPMKKLSASSSTRSIS 874
OY 582 LWNYSMDQAVOSMIN-----TIANGA-----YLEGT----- 609
DB 875 LL-TYANQNHHSINETDIPMAAHIVATQONKKEKAPSPNDQALDHLGSAKRIP 933
OY 610 -----WGSVSNLFFAHDSSGKPIDNMHHRSL-GYLFGIS-----THLDDHSEFLANG 657
DB 934 SKRTESS--SSL--HIPSHNSFSDIHRIKIAHAHAALTAASEKLTPIESTYSSSKA 988
OY 658 OULGKSSDSFISTETTSYIATVQALATSLMKISAOACYNESIHELKTR-----YR 709
DB 989 HNASLSAATLVANKDKLTHAPTPQV-LAPVLEKVLPHRHSVSLQIIRSESEVGOVPFD 1047
OY 710 SFS-KEGGSWHSV-----AVSGEVCAISPIVNSGSLFSSFSIFSKLGFS 755
DB 1048 APSDEKEDLGSSNAPLDVPHGANRSSMDEVNCGSKYDSDGILDGSGQYTD-----SDMS 1100
OY 756 GTQDGFEESSGEIRSFSA 773
DB 1101 DDENSLKDSOSSVLSFSA 1118

```

```

RT Cyclase pathway ".
RL Cell 46:789-796(1987).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE-86220116; PubMed-3011405;
RA Canonis J.H., Kalekine M., Condre B., Garreau H., Boy-Marcotte E.,
RA Jaquet M.;
RT "Characterization, cloning and sequence analysis of the CDC25 gene
RL which controls the cyclic AMP level of Saccharomyces cerevisiae.";
RL EMO J. 5:375-380(1986).
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favello A., Fullon L., Galtung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hallier L., Jier T.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifken L., Riles L., Tatch A., Trevisan E., Vignati D.,
RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.;
RL Submitted (Nov-1994) to the EMBL/Genbank/DBJ databases.
RN (4)
RP DOMAINS.
RX MEDLINE-89181526; PubMed-3070351;
RA Munder T., Mink M., Kuentzel H.;
RT "Domains of the Saccharomyces cerevisiae CDC25 gene controlling
RL mitosis and meiosis.";
RL Mol. Gen. Genet. 214:271-277(1988).
RN (5)
RP FUNCTION.
RX MEDLINE-91203884; PubMed-2017169;
RA Jones S., Vignais M.L., Broach J.R.;
RT "The CDC25 protein of Saccharomyces cerevisiae promotes exchange of
RL guanine nucleotides bound to ras.";
RL Mol. Cell. Biol. 11:2641-2646(1991).
CC -I- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP. THIS
CC PROTEIN POSITIVELY CONTROLS THE LEVEL OF CELLULAR CAMP AT START,
CC THE STAGE AT WHICH THE YEAST CELL DIVISION CYCLE IS TRIGGERED.
CC -I- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X03579; CAA27259.1; -
DR EMBL: M15458; AAA34478.1; -
DR EMBL: U17247; AAB67360.1; -
DR EMBL: U20618; AAB64528.1; -
DR PIR: A26596; RGYC5;
DR HSSP: P29355; ISEM;
DR SGD: S0004301; CDC25;
DR InterPro: IPR000651; -
DR InterPro: IPR001452; -
DR InterPro: IPR001895; -
DR Pfam: PF00617; RASGEF; 1;
DR Pfam: PF00618; RASGEF; 1;
DR Pfam: PF00018; SH3; 1;
DR PRINTS: PR00452; SH3DOMAIN;
DR PROSITE: PS00720; GDS_CDC25; 1;
DR PROSITE: PS50002; SH3; 1;
KW Guanine-nucleotide releasing factor; Cell division; Cell cycle;
KW Mitosis; Transmembrane; SH3 domain;
FT TRANSMEM 1452 1473 POTENTIAL.
FT DOMAIN 58 128 SH3.
FT CONFLICT 497 497 I->Y (IN REF. 2).
FT CONFLICT 954 963 PVGHEPEPKN->LSVTMNSR (IN REF. 2).
SQ SEQUENCE 1589 AA; 179091 MW; CA90053435C85465 CRC64;

```

RESULT 24
 CC25_YEAST STANDARD: PRT: 1589 AA.
 AC P04821.
 DT 13-AUG-1987 (rel. 05, Created)
 DT 01-JAN-1988 (rel. 06, Last sequence update)
 DT 15-JUL-1999 (rel. 38, Last annotation update)
 DE CELL DIVISION CONTROL PROTEIN 25
 GN CDC25 OR CTN1 OR YLR310C OR L2142.6.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-87131091; PubMed-3545497;
 RA Broek D., Toda T., Michelli T., Levin L., Birchmeier C., Zoller M.,
 RA Powers S., Miglier M.;
 RT "The S. cerevisiae CDC25 gene product regulates the RAS/adenylate

RP SEQUENCE FROM N.A.
RC STRAIN-NIZO R5:
RX MEDLINE-93239683: PubMed-8478324:
RA van der Meer J.R., Polman J., Beerlhuizen M.M., Sletzen R.J.,
Kulpers O.P., de Vos W.M.:
RT "Characterization of the Lactococcus lactis nisin A operon genes
nise, encoding a subtilisin-like serine protease involved in
precursor processing, and nlsr, encoding a regulatory protein
involved in nisin biosynthesis."
RL J. Bacteriol. 175:2578-2586(1993).
RN (12)
RP SEQUENCE FROM N.A.
RC STRAIN-6F3:
RX MEDLINE-94213458: PubMed-8161176:
RA Engelke G., Gutowski-Eckel Z., Klesau P., Siegers K.,
Hammelmann M., Entian K.-D.:
RT "Regulation of nisin biosynthesis and immunity in Lactococcus lactis
6F3."
RL Appl. Environ. Microbiol. 60:814-825(1994).
RN (13)
RP 3D-STRUCTURE MODELING:
RX MEDLINE-95357326: PubMed-7630881:
RA Steven R.J., Kollmeier H.S., Kulpers O.P., de Vos W.M.:
RT "Homology modelling of the Lactococcus lactis leader peptidase NisP
and its interaction with the precursor of the lantibiotic nisin."
RL Protein Eng. 8:117-125(1995).
CC -1- FUNCTION: CLEAVES THE LANTIBIOTIC NISIN PRECURSOR PEPTIDE.
CC -1- PATWAY: LAST STEP OF NISIN BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL
(POTENTIAL). BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
-1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
-1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
CC EMBL: L11061: AAA25200.1: -
CC EMBL: X76884: CAA54210.1: -
CC HISSP: P27693: 1AH2.
CC MEROPS: S08.059: -
CC InterPro: IPR000209: -
CC InterPro: IPR001899: -
CC Pfam: PF00082: Peptidase_S8: 1.
CC PRINTS: PR00723: SUBTILISIN.
CC PROSITE: PS00136: SUBTILASE_ASP: 1.
CC PROSITE: PS00137: SUBTILASE_HIS: 1.
CC PROSITE: PS00138: SUBTILASE_SER: FALSE_NEG.
CC PROSITE: PS00139: GRAM_POS_ANCHORING: 1.
CC Hydrolase: Serine protease; Cell wall; Signal; Transmembrane; Zymogen.
KW SIGNAL: 1
FT PROPEP 23 195
FT CHAIN 196 682
FT
FT DOMAIN 196 651
FT TRANSMEM 652 679
FT DOMAIN 680 682
FT ACT_SITE 259 259
FT ACT_SITE 306 306
FT ACT_SITE 512 512
FT DOMAIN 652 657
FT
FT CONFLICT 500 500
FT SEQUENCE 682 AA: 74767 MW: D5F29313F2983EC9 CRC64:
Query Match 2.88: Score 126.5: DB 1: Length 682:

Best Local Similarity 19.9%: Pred. No. 1.3:
Matches 123: Conservative 105: Mismatches 259: Indels 131: Gaps 29:
QY 317 NSQRTSDQGLVNNAYILE-----KDAIISLEARNGLIFEDPIVOESSKESPLPSLQ 371
D 1 HETTNSQQLNININTELINHNNSNAIISSTEGSTDSI-----NLGAOSPAPVKSTT 72
QY 372 ASVTSTPATASPLVQTSA-----NSVITSSERLSSEETPNLTSQDQPIELKSGRLV 428
D 73 RLELDVTGAAKT--LLQTSAYQKEMKVSLOETQVSSEFSKRDSVTN--KEAVPSKDELL 128
QY 429 LMDRAVLSPASLSOD-----PQALLMEAGSLKTSQDLKATLSIPHSID 475
D 129 EOSEVYVSTSSLOKKNILDNKKRANFVYSSPLIKKPPNSDASGV-IDNASPLSTYRK 187
QY 476 TEKSVTHAP---NLSTQKIFLSNGDE--NFENVELLSEQNNIPLLTSLKEOSHLH 529
D 188 AKRVELRQPLKNQKVEAQPPLISNSEKKASVYTSNHFMDQWDMKVVTNNGESYALY 247
QY 530 LPDGNLSSHFQGDWTFPSKDSDEGHSL--IANTPKNYVHPEROSTLVANTLMNTY 586
D 248 QPSKKISVGLIDSG---IMEHPDLSLSLGNVFNKLVPPGGEDENEPDET-----GNP 297
QY 587 SDQAVQSMINTIANG-----GAYLEGTWGSVSNLFYAHDSCKPI--DNMHRNS 635
D 298 SD-----IYDKKHGTVAAGQTTANGNIGVAPGITVINRYF--GENLSKSEVVARA 348
QY 636 LGLVFGISTHSDHS--FCLAGQL--GKSSDSFTSTETTSYIAVQALATSLMK 690
D 349 I-----RAADGKKNIVISAGQYLMISGYDDGNDQYEVLYNKAINTATAGSIY 401
QY 691 ISQAQYNESITHELKRYRFSKKEGSGSHVAVSEGVCSIFIVNGSLFSFSIFSK 750
D 402 VAALGNDLSLIQDNQIMI-NFLKR---FRSIVPQKV-----VDAPSVFED 443
QY 751 LQFSGTQDQFEESGCEIRFSASFFRNISLPICIT--FEKKSQK---TRTYV----- 798
D 444 VIAVGGI-DGY-----GNISDFSNIGADIAIYAPAGTTANFFKYGQDKFVSGYLLKMWLF 498
QY 799 -----YFLGAYIDQLKRDVESGPVYLKNAVSDAPANDSRAYRRLTNQALRLQ 853
D 499 TANTGWQYVYVYVNSFATPKVSGALALVVDKYGKKNP-----NOLKRELLNNSPEVNG-NR 552
QY 854 LNAVSCVLRGOSHSYSLD 871
D 553 VLNIVDLNGKNKAFSLD 570
RESULT 22
ID YJH8-YEAST STANDARD: PRT: 881 AA.
AC P47033:
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHEICAL 89.2 KDA PROTEIN IN SCPI60-SMC3 INTERGENIC REGION.
GN YJL078C OR J1027.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C:
RX MEDLINE-96093911: PubMed-7483841:
RA Miosga T., Schaaff-Gerstenschlaeger I., Chalwatzis N., Baur A.,
Boles E., Fournier C., Schmitt S., Veltan C., Wilhelm N.,
Zimmermann F.K.:
RT "Sequence analysis of a 33.1 kb fragment from the left arm of
RT Saccharomyces cerevisiae chromosome X, including putative proteins
RT with leucine zippers, a fungal Zn(II)-Cys6 binuclear cluster domain
RL Yeast 11:681-689(1995).
RL

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 215035; CA78753.1; -
DR EMBL: X75780; CA53406.1; -
DR EMBL: 228068; CA81905.1; -
DR PIR: B44402; B44402.
DR PIR: S39173; S39173.
DR PIR: S44518; S44518.
DR SGD: 5000151; N0P100.
KW Nuclear protein; Transport. Repeat.
FT DOMAIN 33 571 29 X 6 AA APPROXIMATE REPEATS OF
FT G-L-F-G.
SQ SEQUENCE 959 AA; 99988 MW; D3985F901BBA51 CRC64;

Query Match 2.8%; Score 128; DB 1; Length 959;
Best Local Similarity 20.3%; Pred. No. 1.8;
Matches 143; Conservative 85; Mismatches 257; Indels 218; Gaps 34;

102 NFRPLSTFDSCSKESPITIIHOKNGQLSRNNGSMFCRNHAEGSGAISADAFSLQHN 161
13 NLSEGSNTSSFGGQGS---OQPNLSFGSNNNNNSTNNNAOSGFGFTSAAGSN--SNS 66
162 LPTAFENSSKNGCAIOAOTFSLSRNVPISFARNRADLNGAICCSNLICS--GNVNP 219
67 LF---GNNNNQNNNAF--GQSMGATQN--SPEGLSNNSNNGNTFGSSSMGSGCNTNN 120
220 LFTGNSATN-----GGAICCIDLNTSEKGSISLACNOET----- 255
121 AFNNNSNSTNSPFGFNKPTGCTL--FGSQNNNSAGTSLSFGQSTSTTGTCGNTGSSFG 178
256 ---LF--ASNSAKKKGAIYAKHMYLRN-----GPVSFINNSA-----K 290
179 TGLNGSNIFGAGNNSQNTTGSLEFGNOSSAFGTNNQOGLFGQSOQNTNNAQONQ 238
291 ICGAIAIOTSGSLIILAGEGVLFONNSQRTSDQGLVRNAIYLEKDAILSLFARNQDIL 350
239 LQGS---SFGSKPV--GGSLFGQSN-----NTLGNNTNNRNG--L 272
351 FPDPIQESSKESPLPSL--QASVTSPTPA---TASPLVIOTISANSVIFSSRLSE 404
273 F---GOMSSNGSSNSGLFGONSSNSTQGVFGONNNQMOJNGNNNSLFGKANTFSN 328
405 EKKTPNLISLOQPELKSRLVLRVLAVALSPSLOPOLLINEAGTSLSKSSDKL 464
329 S-----ASGGLFGONNOQSGGLFGQNSOT-----SGSS----- 357
465 ATLSPLHSLDEKSVTIHAPNLISQIFLNSGDNFENVELLSKEO----- 513
358 ---GLFGONNKO-----PNTFTO-----SMTGILFGQN---NNQOQOSTGLFGAKP 399
514 -----NNIPLLLSKQSH-----LILP--DGNLSSHFGY 542
400 AGTTGSLFGGNSSTQPNLSFGTNNVPTSNTOGQNSLFGATKLTMPGCGNTJANQSGS 459
543 GDMTFMKSDSGHSLIANTPKNYP-----HPEROSTLVANT--LWNTYSDMOAVOS 594
460 GNLFGCTKPASTTGLFGNNTASTYPTSTNGILFGNANNASTSTTNGILFGAKPDQSKPA 519
595 MINTIAHGCAVLEFGTQSAVSNLFYAHDSCKPDJMMHRSIAGYLFGISHSIDHNSFCL 654
520 L-----GGGLFGNSNSNSTI-----GONKPVFGTGTGTGLFGATGINS-----SAVG 563
655 AAOQLLKSQSDSTFTTETTSYI--ATVQQLATSLMKISAQA 695
564 STGKLFQGNNTLNVTGONVPPVNNNTONALLCTTAVPSLOQA 606

RESULT 19
YEF3_YEAST
ID YEF3_YEAST

STANDARD: PRT: 956 AA.

AC P32618;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 106.1 KDA PROTEIN IN GLY1-GDA1 INTERGENIC REGION.
CN YEL043W OR SYGP-ORF14.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_Taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Mulligan J.T., Dietrich F.S., Hennessey K.M., Sehl P., Komp C.,
RA Wei Y., Taylor P., Nakahara K., Roberts D., Davis R.W.,
RL Submitted (FEB-1993) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hummel-Smith S.,
RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.,
RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
RN [3]
RP FIBRONECTIN TYPE III DOMAIN.
RX MEDLINE=97148176; PubMed=8994808;
RA Bateman A., Chothia C.;
RT "Fibronectin type III domains in yeast detected by a hidden Markov
RT model. Biol. 6:1544-1546(1996).
RT Cuff. Biol. 6:1544-1546(1996).
CC -I- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U18779; AAB64999.1; -
DR PIR: S30834; S30834.
DR SGD: 50000769; YEL043W.
DR InterPro: IPR001777; -
DR Pfam: PF00041; fn3.1.
KW Hypothetical protein.
FT DOMAIN 35 125 .FIBRONECTIN TYPE-III-
FT SEQUENCE 956 AA; 106132 MW; 3F7B09A0FCA03AF CRC64;

Query Match 2.8%; Score 127.5; DB 1; Length 956;
Best Local Similarity 20.7%; Pred. No. 1.9;
Matches 132; Conservative 91; Mismatches 227; Indels 187; Gaps 31;

88 KEALFR--EKGLDLSIONFRFLSTDCSSKESPSIIHOKNGQL-----SLRNGSM 136
330 KEKFFKLINESNANVANIN---KEIESLONEISKMEESKRLNASKSLITSIVNAVY 385
137 SFCRNHAEGSGAI--SAQDFSLQHN-YLFTAFENSSKNGCAI----- 178
386 ENDKPIASGSLAVLKKLNDFTLEKNFSLNAGEEFLSKLNADSSLSLKMQLSDQEL 445
179 ---QAOTFSLSRNVPISFARNRADLNGAICCSNLICSGNVPLEFTGNS--ATN--- 229
446 EAMWKIDRSNLKLSALEQNFMSLNN-----RNLTKLTVQPKYKNNODSLAATNSNN 500
230 -----GGAICCIDLNTSEKGSLSLACNOETLFSNSAKKKGAIYAKHMYLRN 282
501 SAEKNRSSGSIQPLPSNNMSRTGSDILISNNKRSINNSNADSA-----PLRLHNPV 552

RC STRAIN-C57BL/6 X CBA; TISSUE-Liver;
 MEDLINE-93300511; PubMed-8314577;
 RA Elder B., Laskich D., Gilschler J.;
 RT "Sequence of the murine factor VIII cDNA."
 RL Genomics 16:374-379(1993).
 CC -1- FUNCTION: FACTOR VIII, ALONG WITH CALCIUM AND PHOSPHOLIPID, ACTS
 CC AS A COFACTOR FOR FACTOR IXA WHEN IT CONVERTS FACTOR X TO THE
 CC ACTIVATED FORM, FACTOR XA.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- TISSUE SPECIFICITY: FOUND IN MOST TISSUES.
 CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS. EACH IS COMPOSED OF
 CC 2 PLASTOCYANIN-LIKE REPEATS.
 CC -1- SIMILARITY: STRONG. TO COAGULATION FACTOR V.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: L05573; AAA37385.1; -
 CC PIR: A47004; A47004.
 CC HSSP: P00451; 1CFG.
 CC MCD: MGI:88383; F8.
 CC InterPro: IPR000421; -
 CC InterPro: IPR001117; -
 CC InterPro: IPR00394; Cu-oxidase; 3.
 CC Pfam: PF00754; F5_F8-type_C; 2.
 CC PROSITE: PS00079; MULTICOPPER_OXIDASE1; 3.
 CC PROSITE: PS01285; FA58C.1; 2.
 CC PROSITE: PS01286; FA58C.2; 2.
 CC Blood coagulation: Repeat; Plasma: Acute phase; Calcium
 CC signal; Glycoprotein; Sulfatation.
 CC SIGNAL 1 19
 CC FT CHAIN 20 2319
 CC FT DOMAIN 20 349
 CC FT DOMAIN 20 199
 CC FT DOMAIN 207 349
 CC FT DOMAIN 399 730
 CC FT DOMAIN 399 730
 CC FT DOMAIN 583 730
 CC FT DOMAIN 760 1640
 CC FT DOMAIN 1683 2008
 CC FT DOMAIN 1683 1845
 CC FT DOMAIN 1855 2008
 CC FT DOMAIN 2008 2156
 CC FT DOMAIN 2161 2313
 CC FT SITE 391 392
 CC FT SITE 759 760
 CC FT SITE 1678 1679
 CC FT SITE 1324 1325
 CC FT SITE 1640 1641
 CC FT SITE 1640 1641
 CC FT MOD_RES 367 367
 CC FT MOD_RES 737 737
 CC FT MOD_RES 738 738
 CC FT MOD_RES 742 742
 CC FT MOD_RES 1669 1669
 CC FT MOD_RES 1687 1687
 CC FT DISULFID 173 173
 CC FT DISULFID 547 573
 CC FT DISULFID 1819 1845
 CC FT DISULFID 2008 2156
 CC FT DISULFID 2161 2313
 CC FT CARBOHYD 61 61
 CC FT CARBOHYD 233 233
 CC FT CARBOHYD 259 259
 CC FT CARBOHYD 423 423
 CC FT CARBOHYD 601 601
 CC FT CARBOHYD 880 880

FT CARBOHYD 958 958 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1015 1015 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1022 1022 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1026 1026 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1044 1044 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1046 1046 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1076 1076 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1087 1087 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1136 1136 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1161 1161 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1192 1192 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1255 1255 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1268 1268 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1273 1273 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1274 1274 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1302 1302 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1316 1316 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1340 1340 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1348 1348 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1378 1378 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1797 1797 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2105 2105 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 2319 AA: 266148 MW: FD054DE051DB2A01 CRC64;
 Query Match 2.9% Score 132.5; DB 1; Length 2319;
 Best local similarity 20.2% Pred. No. 3.6;
 Matches 151; Conservative 109; Mismatches 300; Indels 187; Gaps 35;
 QY 46 FEDCT-----MESLEPAL-----CAHASODPELVYLGNSYCNFKLHTDPKALF 92
 DB 775 FKDSITPKNDMEKIPQPEEIAEMLKVOVSYSVDMLMLGSHP-TPHGLFLSDQDEAIY 833
 QY 93 K-----EKGLSTONPFLSFTDCSSKESPSIIHOKNGOLSRNNGSMFC 139
 DB 834 EAIHDSHPMAIDSNESGSKYTQLR-----PESHSEKIVFTPOGLDLSKNSL--- 883
 QY 140 RNHAEGSGAISADAFSIOHNYLFT-----AFEE-----SSKNGCGAI 178
 DB 884 ETTIEVKKKKKLGLOVSSPSNIMTTILSDNLKAFKERTDSSGFPMVHSSKSTTAF 943
 QY 179 QAOTFSLRNWSPISFARNRDLNGAICCSNLICSGVNPFLFTGNSATNGAICISD 238
 DB 944 GKAVSLVGSHPVPLMASENSD-----SNILDS-----TLMYQESLPRONILSIEND 991
 QY 239 LNTSEK--GSLSLACNQLTFLPSNSAKEKGAII---AKHMYLRNGPVPFTNNAKIG 293
 DB 992 RLREKRFHIALLRKDNLFKDNVSLMKTNTYHSTNEXLHTESPSTSIENSTTDOD 1051
 QY 294 AI-----AIOGSGSILAGEGSYLFONNSORTSDGLVRNAILYLEKDAIIS-SLEFARG 347
 DB 1052 AILKVNSEIO---EYALALHDGTLAKNST-----YLRNMLMARTSTKTK 1095
 QY 348 DILF--FDPIVOESSKESPLPSLSQASVTPPTATASPLVIOTSANRVSIFSSERLSE 404
 DB 1096 DLFHREKDEDPIDODEENTIMPSKMLFLS-----ESSNMFKKTGNNSL-----NSE 1142
 QY 405 EELTPNLTSLQOQPIELKSGRLVLDRAVLGAP---SLSDOPQALLIMEAGTSLKTSDD 461
 DB 1143 QEISPPQOLVYLM-----FKRYVKNQSFLEKKNVYIEQDGFRTNLIKDMAFPHNMS 1194
 QY 462 LKATISILHS---LDTEKSV--TIHAPNLSTOKIFLSN---SGDENFENVELLSKE 512
 DB 1195 IFLTLIS-NVHENGRIKNOGKNIQIELEKALIEKVVLPQVHEATGSKNFKLIDILIGTR 1253
 QY 513 QN-----NIPPLTLSEQSHLHPDGNLSSHFEGYQGDWTFMSKDDDEGISLA 560
 DB 1254 QNISLEYVHPVLPQNTITSINNSTNTQIHM-----EHFFKRR-----KDETNSEGLV 1301
 QY 561 NMTPKVYVHPREKOSTLVNLTMTNTYSDMAVQSMINTIAHGCAVLFGTMGSAVSLFPA 620
 DB 1302 NKTR-----WKN-----TPSQKNITTSKRALGOFRLSTOW---LKTIKNS 1342
 QY 621 HDSSGKPIDMMHRSILGYLFGISTHSLDHSFCLAAGOLGKSSDSFISTETTSYIATV 680
 DB 1342 NKTR-----WKN-----TPSQKNITTSKRALGOFRLSTOW---LKTIKNS 1342


```

Oy 452 AGTSLKTSDDLKATLTIPLHSLDT-EKSYTHA---PULSTOKIFLNSGDNFENV 506
Db 824 SGVEIEST-----TSFSFHETSTASTVOISSQFVTPSPSITVAPRSTGLNSTEST 878
Oy 507 ELLSKEDONNIPLLTLAKFOSHLHPDGNLSS-----HGCYGDWTFMSKXDEG 555
Db 879 N-SSEK-----TWSSSNASVMPSSASPTCKVTSOETSSEGFRRDRTYYRMTSE- 929
Oy 556 HSLIANTPKNYVPHPEROSTLV-----ANTLNTYSQDMQAVQSMINTIAHGCAVLEGT 609
Db 930 -----IPSTNEQTLITLVSSCESNSCNTVS--SAVYSTAFTTINCITTEYTT 975
Oy 610 WGSANVLFLYAHDSGKPIDNMHHRSLGLYLFISTHSLDHSFC-LAAGQLGKS---SD 665
Db 976 W-----CPLSLATELTLYTSKLESE 993
Oy 666 SFTTSTETSYIATVQALSTLKKISNQACYNESIHELKTKRYSFSGEFGSMHSAVS 725
Db 994 EKTLLITVTSCEGVCSTASPAIVSTATAVNDV-----TWISTWFO--AINKLAVS 1046
Oy 726 GEV-----CASIPVNSGGLFSSFSJFSKLQGSFGTQDGFEESSGCEIRSFSA 773
Db 1047 SDIENSASKASFVEEAEFTKISIRNNN-FVPTSGTSTIEHTTTTSNASENSDVNSSEA 1105
Oy 774 SDFRNISLPIGITEFKKSQKT 794
Db 1106 VSKSVTNPLVISVSOOPRGT 1126

RESULT 10
Y144_HUMAN STANDARD: PRT: 983 AA.
AC 014157;
AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN KIAA0144.
GN KIAA0144.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96127530; PubMed=8590280;
RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. IV.
RT The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by
RT analysis of cDNA clones from human cell line KC-1.";
RL DNA Res. 2:167-174(1993).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D63478; BAA09765.1;
DR InterPro: IPR000449;
DR Pfam: PF00627; UBA: 1.
DR Hypothetical protein.
SQ SEQUENCE 983 AA: 103930 MW: 50579GCFa293EFA68 CMC64;

Query Match 3.1%; Score 139.5; DB 1; Length 983;
Best Local Similarity 23.0%; Pred. No. 0.35;
Matches 141; Conservative 64; Mismatches 226; Indels 183; Gaps 30.
237 SDLNNTSEKSGSLIACNQETLFPASNS-----AKKGAITAKHHVLRNYRNPYSPFINNSAKI 291

```

or send an email to license@isb-sib.ch.

CC -----
 CC EMBL: M77354; AAA34798.1;
 DR EMBL: 272799; CAA96997.1;
 DR PIR: S25370; S25370.
 DR SCD: S0003246; MSB2.
 KW Transmembrane: Glycoprotein; Repeat.
 FT DOMAIN 698 816 7 x 17 AA TANDEM REPEATS.
 FT REPEAT 698 816 1.
 FT REPEAT 715 731 2.
 FT REPEAT 732 748 3.
 FT REPEAT 749 765 4.
 FT REPEAT 766 782 5.
 FT REPEAT 783 799 6.
 FT REPEAT 800 816 7.
 SO SEQUENCE 1306 AA; 133114 MW; 675D5D984D5CA4A6D CRC64;

Query Match 3.3%; Score 149.5; DB 1; Length 1306;
 Best Local Similarity 21.1%; Pred. No. 0.12;
 Matches 161; Conservative 106; Mismatches 347; Indels 149; Gaps 33;

OY 107 STTDCSSKSSSIHQ-----KNGQLSLNNGSMSCRNIADESGGAISDAFSLQH 159
 DB 496 SSSDVAS-TSAPSVYSSSFTYSLAQGSSMTNPSSSTIVYSSSTGSSREBSASTAS--- 551
 OY 160 NYLTAFAEENSKNGCAIOAOTFLSRNVSPISFARNRADLNGAICCNLCGNVP 219
 DB 552 ----ATLSSSTTYMAGNIQSDPSTSLLS-----ESQATSTANVLASSV---STSP 599
 OY 220 IFFTGSAATNGAICCTSDLNTSEK-SLSLACNOETLPASNS---AKEGAIYKKNV 215
 DB 600 YTAGAGATSEASLISSTSAETSOVSQSTALQTSFSSSTEGSESTSSV 659
 OY 276 LKRVGVSTIN-----NSAKIGAIAGSGSILAGESVLEPQNNSQRTSDQ 324
 DB 660 L-VQWSSSISSEPSQTTOMNSASSSSOYTTISSTGLSQVS-DTSVSTTSSSVSOV 717
 OY 325 GAVRNAIYLEKDAIISLEARNGLDFFDPYVDESSKESPEPSLQASVTSPT---AT 381
 DB 718 SDTPVSTYTTSSSVSOVDTPVSTYTTSSSVSOVDTPVSTYTTSSSVSOVDTPVSTY 777
 OY 382 ASPLVITQTSANRSVIFSSER--LSEBEKTP-----DNLTSQLQPIELKSGRLVK 430
 DB 778 SSSSVSOVS-DTSVPTSSRSRSVSOVDTPVSTSSRSVSQTSSTSLQPTTSSORTIS 836
 OY 431 DRAVLS-APSLQDPQALLMEAGT-----SLKTSIDLKATLSTI-----PLHSLDEKS 479
 DB 837 THGALSSSSSVSOQASEITSSINATASEVHSIOTTAATOSTTSLFTDANSSASAPLEVA 896
 OY 480 VTIHAPLISLOKIFLSNGDENFENELLSKEONNIPILT-LSKEOSHLLPDLGNLSH 538
 DB 897 TSITPTSKKSSLLTPTS-----TSSLQVATNTNVQTSITLTESTTLYLEPTTSSS 948
 OY 539 FGVOGDWTFKWKOSDECHSLANNTKKNVYPHPRQSTLVANTLMNTYSDMAQVSMINT 598
 DB 949 -----TFSLVNTSSDMNM-----WIPTELITQAEASTASTVGGT-----QTM--T 988
 OY 599 IAHGATLFGTWSAVNLFVAHDSGKP-IDNMHRSLSLGLFGISTHSLDDHFCILAAG 657
 DB 989 LPH-----AIAAATQVPEPECTTLITICFKALINFEVVSERK-----SSA 1029
 OY 658 QLLKSSDSFT-----STETSYIATVOALATSLMKISQAQYNEIHELK---TK 707
 DB 1030 QIFGLYDALNTPKKNFTNTITVLOIYPLDDSLNLYLVAEVEFPAETIELSLNLTNS 1089
 OY 708 YRSFSKEGFGMSVAVSGEVCASIPV-----SNGSLFSSTFSFKLQGGSG----- 756
 DB 1090 SSARFYTQMGAKMAAM--VDSSIPILGLHDNSNNGSGSDSSSSNSNSGSGSGSN 1147
 OY 757 TQDFEESGE-----IRFSASSFRNISLPITGTFEKKSK 793
 DB 1148 SNGSVSSSGNSYODACTLEYSSKSNMVS-----TSKSKKK 1185

RESULT 7
 Y109-YEAST
 AC P40442:
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HYPOTHETICAL 99.7 KDA PROTEIN IN SDLI 5' REGION PRECURSOR.
 GN Y1169C OR Y19402.07C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
 RA Churcher C.M., Connor R., Copley T., Dear S., Devlin K., Fraser A.,
 RA Gentles S., Hamlyn N., Horneill T.S., Hunt S., Jagels K., Jones M.,
 RA Louis E., Lye G., Moule S., Moule T., Odeil C., Pearson D.,
 RA Rajandream M.A., Riles L., Rowley N., Skellon J., Smith V.,
 RA Walsh S.V., Whitehead S.;
 RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 246921; CAA7023.1;
 DR SCD: S0001431; Y1169C.
 KW Hypothetical protein; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 995
 FT CARBOHYD 28 28
 FT CARBOHYD 35 35
 FT CARBOHYD 468 468
 FT CARBOHYD 664 664
 SO SEQUENCE 995 AA; 99735 MW; F63E287A03F137EC CRC64;

Query Match 3.3%; Score 148.5; DB 1; Length 995;
 Best Local Similarity 19.8%; Pred. No. 0.095;
 Matches 130; Conservative 112; Mismatches 278; Indels 135; Gaps 23;

OY 181 QTSLSRNVPISFARNRADLNGAICCNLCGNVNPFFNGNSTNGAICCTSDLN 240
 DB 24 QVYNSNTSISNSSSTSVSSSGSVSISSIAE-----TSSATD--ILSITQSA 73
 OY 241 TSEKGSILACNOETLPASNSAKEGAIYAKHVLRYNGVPFINSKIGAIAGIAG 300
 DB 74 SSTSISVSSVSSSVSSSVSSSVSSSS-----VSVSSSVSSSVSSSVSSSVSS 120
 OY 301 GSLSILAGESVLEPQNNSQRTSDQGLVRNAIYLEKDA-----ILSLEARNGLDFFDPY 357
 DB 121 SSVSQASASTSDVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSS 180
 OY 358 ESSKESPEPSLQASVTSPTPATASPLVITQTSANRSVIFSEELSEBEKTPDLTSLQ 417
 DB 181 QASASTSDVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSS 236
 OY 418 OPTLKSGLRLVKDRAVLISAP-SLSDPQALLMEAGTSLKTSIDLKATLSTIPLHSLDT 476
 DB 237 QASASTSGVSSSGSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSS 296
 OY 477 EKVY-----TIHAPNLSTIO-----KIFL-----SNGSDENFYE 504
 DB 297 SNLSLSSDGTIYIPPTTISGDLTLTKVIAEGVVAAGAKLTLLDGDKYFSFADLVYGG 356


```

QY 179 Q-AOTFSLSRNVSPISFARN-----RADLNGAICCC-SNLCGNNVPLEFTGNSATNG 230
D 184 STAFFVSEVSENOGFLFMDNICIOTNAGKAIYAGTSNFSNENMCD-LFFINNACGAG 242
QY 231 GA-----IC-----CISDLNT-----SEKGSISLACNOST-----LEFAS 259
D 243 GAIFSPICSLTGNRCNIVFYNNRCFKNVETASSSEASDGAIRKVTTRLDVTNRCGRIFPSD 302
QY 260 NSAREKGAIAKIMVLYRNGVPSPFINNSA-KIGGAIAIOGSGSILAGESVLFQCN- 316
D 303 NITKNYGAIAIAPVTLVDNGPTFTINNANKKGAIIYIDGTSNKISADNHAIFENMI 362
QY 317 -----NSORTSDGCLVNAIYLEK-----DAILSLEARNGDILFEDPI-----VOESS 360
D 363 VTNTNAGTSTSNAPPRRNATIVASSGEILLGAGSSON-LIFEDPIEVSNAGVSYSF 420
QY 361 SKESPFLPSLOASVTSPPTATASPLVIOTSANRSYIFSSERISEEKPDPNLTSLQLOPI 420
D 421 NKEAD-----QTG-----SVFSGATVNSADFHQRNLQTKTPAPL 455
QY 421 ELKSGRIYKIDRAVLSAPLSQ-----DPOALLIMEAGTS----- 455
D 456 TLSNCFCLIEDHIAOLTVNRFTOTGVSISLNGAVLSCYKNGTGDASANSATILKHIGLNL 515
QY 456 ---LKTSSDLKATLISIPLSLDTREKSVTIHAPNLSTQIFLSNGDENFYENVEL---L 509
D 516 SSLKSGAIEPIPLWE-PTNNSNNTADTAFTSLDVKLSIDYGNSPYSTDLTHAL 574
QY 510 SKEONNIPILLTSKEOSHLHPDGNLS-----SHFGYGDWTFEKKDSE----- 554
D 575 SSO-----PMLSTIS-EASDNQLOSENI DFSGLNVPYHGWGLTWMAKQDDEPPASSAT 629
QY 555 -----GSLIAMPTEKKNYVPHPEROSTLVANTLMTYSDMOAVQSMINT--IANGG 603
D 630 TDPQKANKRHRILLTLWLPAGYVPSPKHRSPLIANTLMGNN--LATESLKNSAELTPSG 687
QY 604 AVLFGTWSAVSNLFYAHDSCKPRIDNMHRSIGYLFISTHSLDHSFCLAQOLGKS 663
D 688 HRPWGTGGGLGMVYQDPRENNP--GFHRSRSGYSGAMTAG--QTRHESLAKFSQTYTKL 743
QY 664 SDSFTSTETTSIATVQAOALATSLMK-----ISAOCYNESIHLELTKRYSFSEKEG 716
D 744 NENVY-AKNVNSAANYSCQCEMLFSLQEGFLYKLVGLYSTGDHNCHEFTYOGNLTISO-- 800
QY 717 GSHSVAVSGEVCASIPYVNSG-----GLFSFSIFSKLOGFSGTQDGEES 764
D 801 GTRSGOTMGCAVFEDLPKRFSGTHILLTAPFLGALGIYSSLSHFTVGAYP----- 851
QY 765 SGEIRFSFAS--FRNISLPIGI--TEKKSQKTRTYVYFLGAYIQDLKRDVESCPLYL 821
D 852 -----RSSTRTPLINVLPIGVKSGFEMNATHRPQAMTVEL-AVOPVLYROEPGIAOQLLA 906
QY 822 NAVSWDAPMANLDSRAYMERLINO 845
D 907 SKGIWFGSGSPSSRHAMSXYKISOO 930

```

RESULT 2

```

ID OMPF_CHLTR STANDARD: PRT: 1034 AA.
AC P38008; 084878;
DT 01-OCT-1994 (Rel. 30, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PUTATIVE OUTER MEMBRANE PROTEIN F PRECURSOR.
CN PMPF OR CT870.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxId=813;
RN 11;
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/CX;
RX MEDLINE=99000809; PubMed=9784136;

```

```

RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
RN 121
RP SEQUENCE OF 26-35.
RC STRAIN=L2/434/B0;
RA Blui L., Santucci A., Magi B., Marzocchi B., Sanchez-Campillo M.,
RA Comanducci M., Christensen G., Birkelund S., Vitreou E., Ratti G.,
RA Pallini V.;
RA Submitted (SEP-1994) to the SWISS-PROT data bank.
CC - SUBCELLULAR LOCATION: CELL WALL, SURFACE (ELEMENTARY BODIES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC
CC DR EMBL: AE001360; AAC68468.1; -
KW Outer membrane; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1034 PUTATIVE OUTER MEMBRANE PROTEIN F.
FT CONFLICT 34 35 RR -> FH (IN REF. 2).
SQ SEQUENCE 1034 AA; 112392 MW; 445FFC435D463AE7 CRC64;

```

```

Query Match 9.9%; Score 451; DB 1; Length 1034;
Best Local Similarity 22.3%; Pred. No. 5.8e-21;
Matches 223; Conservative 146; Mismatches 367; Indels 264; Gaps 41;

```

```

QY 63 SDDIPLYVGNISYSCWFSKIAHITDPEALFKKGDLSION--FRPLSTFD-----C 111
D 105 ASDNANMLFCNNYC-----THNKGGAIRSGGPTRFLNNDVLYFNINISA 149
QY 112 SSKRESSPSIHKKN-----GOLSRNNGSMEFCNHAEGSGAISA-- 154
D 150 GAVYGTGDHNEKRNKGALYAFITLTGNKRTLAFINNMGDGGGALISADTOISITDVTGK 209
QY 155 --FSLQHNILFTAFENSSGNGCAL--QAOTFSLSRNVSPISFARNRDLNGALICCSN 210
D 210 ILENNHTLNHPYTOAENNARGAICSRDLCSISNNSGPYFNNNGG-KGGAISAR 268
QY 211 LKICGVNPLPFTGN-----SATNGAICISDLN-TSEKGSISLACNOETLFPASN 260
D 269 CVIDNNKERTIFENNSSLGSSOSSASNGAIOGTGQFLRNKKGSI-----YFDSN 320
QY 261 SAREKGAIAKIMVLYRNGVPSPFINNSAKIGALAIQSGSLS--LAGEGVLFOGNS 318
D 321 TATHGCAIINGYIDIRDNGPYVFLNNSAAMGAFLNLSKPRSATNYIHGTGDIVNNNN 380
QY 319 QRTSDGIV--RNAIYLEKAL--LSLEARNGDILFEDPI-----VOESSKESPLPS 368
D 381 VFTLDGNLGRKRLFRHNNNEITPYTSLGAKKDRIFYFDLFQMEVKNKTSNNPSPPT 440
QY 369 SLOASVTPPTATASPLVIOTSANRSYIFSSERISEEKT-----PDNLTSLQLOPIELKS 424
D 441 SRNITVNP-----ETFSGCAVVFYNGMSDRIPLMKENHYIK--EAPTLKF 488
QY 425 GRVLKRAVLSAPSL--SDPOALLIMEAGTSIKTS--SDKLATLISPLHSL----- 474
D 489 GTALIEDDAELEIFNIPFTONPTSLALGSGATLTGKHKGLNTNGVILPITILEGKS 548
QY 475 -----DTEKSVTHAPN--LSIQKIFLSNSG-----DENFYENVELSKQONNIP 517
D 549 PCIRIVNPQDNTONTGTGOTPPSSSTSSISTPMIIRNGSLIVDENVESVYDSMDLSRKA 608
QY 518 ILTLTSKEOSHLHPDGNLSS-----HFGYGDWTFEWS----- 549

```


OY 521 LKESQSHL-----HLPDGNLSHFYOGDMTFSMK-----DSDEGSLIANMTP 564
DB 610 IDVDLHLGAPRANAPRDLTLGESSKTYGOSMTLQMEPRPAPPNQNSNMLKASMTK 669
OY 565 KNYVHPERQSTLVANTLMTNTYSDMAVOAMINTIAHGAVLFTGWSAVSNLFYAHDS 624
DB 670 TGYNGCPERVASLVSNSLWGSILDVRSASHASIAOASIDGRACRGIMISNFFY-HQDD 728
OY 635 GKPIDNMHRSGLYFGISTHSLDHSFCLAGQLGKSSDSFISTE-----TSYAT 679
DB 729 ALG-OGYRHISGYSIGANSY-FGSSMFLAETETFGGRKSDVVCORSNDHTCVGSVLYST 786
OY 660 VOALATSLM-KISAQACNESIHELKTRYSFSGEGSMHSAVSGFCAPIVSNG 738
DB 767 ROALGSLFLGDAFRASTYGFNGQHKTSY-TFAESNVRMDNVCVGEVGLPIMLA 845
OY 739 SGLF-SSEFSIFSKLOGFGOTODGFESSGEIRSFASFRNISLPIGITEFKKSOKRTY 797
DB 846 SKLYINELRPFVQAEFAVAEHESFTEGDAQAREFKSGHLMNLSIPVGYFDRCSKHPMK 905
OY 798 YFFLCATYDLDKRDVESGPVLLKNAVSDAMANDRAHYFRLTNORAL-HRLQTLN 856
DB 906 YSFMGAYICDAVRSISGTETTLTSHKETTTDAFHILRHGVYVRSMYASLTGNIYVGH 965
OY 857 VSCVLRGOSHSYSLDICTYRF 878
DB 966 GKYEYRDASRGYGLSIGSKIRF 987

RESULT 22

084879 PRELIMINARY: PRT: 1013 AA.

AC 084879: 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE PUTATIVE OUTER MEMBRANE PROTEIN G.
GN PMPG.
OS Chlamydia trachomatis.
OC Bacteria: Chlamydiales; Chlamydiaceae: Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D/UW-3/CX:
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 287:754-759(1998).
DR EMBL: AE01360; AAC68469.1;
SQ SEQUENCE 1013 AA: 107366 MW: F0927743C0A651DD CRC64;

Query Match 16.7% Score 758; DB 2; Length 1013;
Best Local Similarity 26.2% Pred. NO. 19e-45;
Matches 263; Conservative 139; Mismatches 399; Indels 208; Gaps 34;

OY 26 DPLGETA-----LTKRPMHVVCTFEEDCTMESLPALCAHNSODPLVLGNSYCMFVS 80
DB 53 DPGSTVFSAGELTLKLNLSIALPLSCFGLN-----GSFVLGSGHSLTFRE 101
OY 81 KLHTDPKAL--FKKGLSTONFRFLSFTDCS-----KSSPS- 119
DB 102 NRTSTNGAALDSANSCTFTIEGKELSFNSCNLSLAVLPATTNNGSOTPTTSTPSN 161
OY 120 -IIFHOKNGQLSRNNGSMFCRHHAGSGGASADAFSLQHNLYLTAPEHNSKKGCGAI 178
DB 162 GTTYSKT-DLLLNNEKFFYSNLSVSGDGAIDAKSLTVQGISKLCVFOENTAOADGAC 220
OY 179 QAOT-FSLSRNVSPISFARNRADLNGAIC----- 208

DB 221 QVNTSFSAMANEAPIAFIANVAGVRCGIAAIVODGQCVSSSTEDPVSFSTRNATVEF 280
OY 209 -----SNLIGSNV-----NPLFFIGNSATN-----GGA 233
DB 261 DGNVAVGGCIYSGNVAFNLNGKTLFLNNVASPYIAAEOPNTNGQASNTSDNKGCGAI 340
OY 234 CC---ISDLNTSEKSLSLACNOETLFPASNSAKEGCAIYAKHMLRPNCPVSFIINSK 290
DB 341 FCKNAQAAGSNNSGVYFDEGEVYFFESSNVAAGGCAIYAKKLVSANCGVQFLGNLAN 400
OY 291 ICGAIAIAGSGSLSLAAGEVSLFQNNQSORTSDQGLV-NALYLEKDAI-----LSSL 342
DB 401 DCGAIYLGESGSLSLADYGDILFDGNLKRPAKANAADVNGVTVSSQAISMGGKITTL 460
OY 343 EARNG-DILFPPIYOESSKESPLPSLSQASVTSPTATASPLYIOTSANRS-VITSS 399
DB 461 RAKAGHQLFNDPI-EMANGNQPQSS-----EPLKINDGEYTGDIYA- 505
OY 400 ERLSEEEKTPDRLTSQLQPIELKSGRLVKDRAVLSAPSLSDPQALLMEAGTSL--- 456
DB 506 -----NGNSTLYQNVITIEGRIVLREKALVSNLSQOTGS-LYMEAGSLDFV 553
OY 457 -----KTSDDLKATLSIPLHSL-----DTEKSVTI 482
DB 554 TPQPOOPPAANOILITLSNLHLSLILANNAVTPNPTNPADSHPAIIGSTTAGSVTI 613
OY 463 HAPNLSIOKIFLNSGDENFYENVELLSKEQN---NIPLLTSLSEQSHLHPQNSLSH 538.
DB 614 SGP-----TFEEDL-DQVAVDRLDGLSNOKIDVLKQLQLOPSANAPSDTLGNEPK 666
OY 539 FGYGDMTFWS--KDSDEG-HSLIANMTPKNVHPEROSTLVANTNTYSDMAVOAM 595
DB 667 YGYGSMKLANPNTANNNGPYTLKATWTKTGYNCPERVASLVSNSLIDISALISA 726
OY 596 INTIAHGAYLFTGWSAVSNLFYAHDSGKPIDNMHRSGLYFGISTHSLDHSFCLA 655
DB 727 IQASVDGRSYCGLWVSGVSNFFY-HDRDALG-OGYRYISGYSIGANSY-FGSSMFLA 783
OY 656 AGQLLGKSSDSFISTE-----TSYIATVOALATSLM-KISAQACNESIHELKTYR 709
DB 784 FTEVFGRSKDYVYCRSHNHACIGSVYLSLTKQALGCSYLGDAFIRASYGFGNQHKTS- 842
OY 710 SFSKGEFGSMHSAVAVGECASPIVNSGCLF-SFSIFSKLOGFGSGTODGFESSGEI 768
DB 843 TFAESDVAMDNCLVGLGVGLPIVITPSKLYLDELRFVQAEFSYADHESFTEGQA 902
OY 769 RFSFASFRNISLPIGITEFKKSOKTRYTYFLGAYIIDLKRDVESGPVLLKNAVSDA 828
DB 903 RAFRSGHLMNLSVPYGVKFRDCCSTHPNKSIFPMGAYICDAYRTISGTOTTLTSHQETWT 962
OY 829 PMANDSRAMFRLTNORAL-HRLQTLNVSCVLGQSHSYSLDICTYRF 878
DB 963 DAFHILRHGVIVRGSMYASLTGNIYVGHGREYRDTSGYGLSAGSKYRF 1013

RESULT 23

092965 PRELIMINARY: PRT: 922 AA.

AC 092965: 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY PMP_1 PROTEIN.
GN PMP_1.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria: Chlamydiales; Chlamydiaceae: Chlamydia.
OX NCBI_TaxID=83358;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CW1029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kaiman S., Mitchell W., Marathe R., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;

641 G1STRDSDHSEFLAAGLQGLKSSDPSFITSTETSYATVQAOL----- 684
 575 GYVAKTPEDDLSFAFCOLLGKDKDYLVSKNNANIYASLIYQH1SYSAMONLLQNTIG 634
 685 ATSLMKISAQACYNESIHELTKY-----RSFSKEGFGSMHSAVAGVCASIPIVSN 737
 635 AEAFLVLAQALTYCHASNDKMTNMTTYAPKRTYAEIKGDMGNDGFCVELGATVPIOTE 694
 738 GSGLFSSSISIKLOGFGTODGFEE-SSGELRSASASFRISIPITGITEPKKSQKRT 796
 695 SSLSDMISPLKFLVHTHODDEFKENSDDGRYFESSNLNLSPIGIKFERFANNDTA 754
 797 YYFLGAYTODLKRDESGP-----VVLKNAVSWDAPMANLDSRAVMEPLTNOAL-HR 850
 755 SYHTAAATSPDI---VRNPDCITSLVSPDSAVVVTKANLARSAPMLQAGNYLSHN 811
 851 LQTLNASCVLRGQSHSYSLDGTTRF 878
 812 IEIFSQGFELRGSSRTYVNDLGSKIQF 839

RESULT 19
 092898 ID 092898 PRELIMINARY: PRT: 936 AA.
 AC 092898;
 DT 01-MAY-1999 (TReMBLrel. 10, Created)
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
 DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
 DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY.
 GN PMP_7.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CWL029;
 RX MEDLINE=99206606; PubMed=10192388;
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.,
 "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."
 RL Nat. Genet. 21:385-389(1999).
 DR EMBL: AE001627; AAD18589.1;
 SQ SEQUENCE 936 AA; 100079 MW; 8881D78A5D3D194EC CRC64;

Query Match 17.1% Score 773; DB 2; Length 936;
 Best Local Similarity 28.0%; Pred. No. 1,4e-46;
 Matches 236; Conservative 143; Mismatches 378; Indels 86; Gaps 22;

97 DLSIONFFFLSFTDCSSKSSPS--IHQKNGOLSLRNGSMSPCRNHAEGSGAISADA 154
 119 NLTFNDFRSLIISCPILLSPTGCGALKSVGNLSLTGNSQIIFQNSSDNGVYINKN 178
 155 FSLQHNLYFTAFENSS-KNGGAIQAQTFSLSRVSP--ISFARNRDLNGAICCSNL 211
 179 FLSTGTSOFASFSRQAFTGKGGVYATGTTIENSPIVFSQNLKGGSGALYSTD- 237
 212 ICSGVN-PLFTFGNSA-----TNGAICCSIDLNTSEKGSLSLACNOETLFASSAEEK 265
 238 NCSITDNQVIFDGNSAEAAQAQGAIC---TTTDR-TVTLGNANLSFTNTALTY 292
 266 GCAIYAKHMYLRNGPVFINNSAKI-----GGAIAISGSGSLILAGEGVLEFQNSQ 319
 293 GCAISGLKLVTSAGGPTLFQSNISGSSAGGGGAGINIASAGELALASTSGDITF-NMQ 351
 330 RTSQGLVNAIYLEKDAILSLEARNG-DLFPDPIVOESSKESPLSSIQASVTPPT 378
 352 VTNSTSTIRANINIIDTKVTSIRATQSIYFDPIINPGTAATDNLNLADANS-- 409
 379 PATASPLVITRANSRVFSSERLSEEEKT-PDNLTSOLOPIELKSGRLVAKDAVLSA 437
 410 -----EIEYGAIVFSEKUSPTKAIANVTSTIRPPAVILARGDLVLRDGVITVF 460

438 PLSDDPQALLMEAGTSLKT-SSDLKLATISLPHSLDTEKSVYIHA-----PNLSIOKI 492
 461 KDLTSPGSRILMDGGITLSAKEANLNLNGLAVNLSLDGNKKAALKTEADKNSISLGT 520
 493 FLNSGDNFENVELLSKEQNNIPLLTSLKEOSHLLPDNLS-----SHFGYOGD 544
 521 IALIDTEGSPFENHL--KSASTYPLLELTAGANGTTLGALSTLTLOEPETHYQYGN 578
 545 WFSKDKDDEGSLIANTPKNYVHPEROSTLVANTLMTYSDMQAVQSMINTANGA 604
 579 WQSWANATSSKIGSINRTGCIYIPSPERKSNPLNSLWGNFIDIRISNOIETKSSGEP 638
 605 YLFCWGSVANSILFYAHDSCKPIQNMHHRSLGYLFGISTSLDHSCLAGOLGSS 664
 639 FERELMWSGINAFYRDSMPTR--HGFRIHSGGALGITTATPAEDQLTFACOLFADR 696
 665 DSFITSTETSYATVQAOLATSLMKIS-----AQACYN 698
 697 NHITGKNHCDIYGASLYFHHTGCLPDIANFLMGKATRAPWVLSLSEISQIIPLSFPAKPSYL 756
 699 ESHELTKY--RSFSKGFSGSMHSAVAGVCASIPIVNSGGLFSSFSIFSKLOGFGSG 756
 757 HTDNHMKTYTYTDSNIIK---GSMRNDACADLGSALPFIIVPYLLKEVEPFVKQYTA 813
 757 TODGFESSGFIKRSASFRNISLPIGITEPKKSQKRTTYFLGAYTODLKRDESGP 816
 814 HQODFERHAGRAFNKSELINVEIPGVTFERDSKSEKGYDILMTLIDAYRNPKCQ 873
 817 VVLKNAVSWDAPMANLDSRAVMEPLTNOALH-RLQTLNASCVLRGQSHSYSLDGT 875
 874 TSLIASDANMMAYGTNLARQGFSAVAAHFOVNPHEIFQGFARFVRSNNYNTNLGSK 933
 876 YRF 878
 934 FCF 936

RESULT 20
 093542 ID 093542 PRELIMINARY: PRT: 936 AA.
 AC 093542;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
 DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).
 GN PMP_7 OR CP0308.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shiba M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.,
 "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA."
 RL Nucleic Acids Res. 28:2311-2314(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 White O., Hickey E.K., Peterson J., Umayam L.A., Ullrich T.,
 Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
 Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J.,
 McClarty G., Salzberg S.L., Eisen J., Fraser C.M.,
 "Genome sequences of Chlamydia trachomatis Morn and Chlamydia pneumoniae AR39."
 RL Nucleic Acids Res. 28:1397-1406(2000).
 DR EMBL: AP002546; BAA98653.1;
 DR EMBL: AE002193; AAF38165.1;

```

OY 708 YRSFSEKGSMSHVAAGVCAPIVSNCGSL-FSSSFISFKLOGFCGQDGEESGC 766
DB 824 YTKLPK-GRCSNSNMCWGELEGELPILTSRLNKQIIPVKAQVAYATGGIOENTP 882
OY 767 FLSFSSASSFRNISLPICITTFPEKKSOKTRTYVYFLGAYIQDLKRDVEGCPVLLKNAVSM 826
DB 883 EGRIGHGHLNVAAPVGVCFKNSHNRPDPFYITIAVAPDVYRHNHPDODDTLLPINCATW 942
OY 827 DAPMANLDSRAVMEFLTNORALHR-LQTLNVCYLRGOSHYSIDLGTTFYF 878
DB 943 TSGNNLTRSTLLVQASSHTSYNDVLEIFGHCGCDIRRTSRQYTLIDIGSKLRF 995

RESULT 16
P71133 PRELIMINARY: PRT: 846 AA.
AC P71133:
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE POMP91B PRECOURSOR.
OS Chlamydia psittaci.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RX MEDLINE=96406378; PubMed=8810511;
RA Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;
RT "Molecular cloning and characterization of the genes coding for the
RT highly immunogenic cluster of 90-kilodalton envelope proteins from the
RT Chlamydia psittaci subtype that causes abortion in sheep.";
RL Infect. Immun. 66:1317-1324(1998).
DR EMBL; U65943; AAC15923.1;
KW Signal.
FT CHAIN 1 16 POTENTIAL.
FT SIGNAL 17 846 POMP91B.
SO SEQUENCE 846 AA; 90834 MM; 4CDC31DC03C2964E CRC64;

```

Query Match 17.4%; Score 790.5; DB 2; Length 846;
 Best Local Similarity 26.2%; Pred. No. 7e-48;
 Matches 264; Conservative 141; Mismatches 363; Indels 169; Gaps 35;

```

OY 11 LCAAILSTAVLFCQDPLGELTALTKNPN-HVYCTFFE--DCTMESLPFA--LC-AHAS 63
DB 10 ISSLLASNSLSFAQ-VNTEITLTSDSYNGVNTSDPEFKETTSQAIYVCEGNVCISVAG 68
OY 64 QDDPLVGNSTCMVFSKLIHTDKRELFKEKGDLSIONFR-----LSFTDCSKESKES 117
DB 69 KDSPL-----NSKC-----FSE-----TTEMLSFIGNGYTLCPDNITTSOSH 105
OY 118 PSIIHQKQGLSLRNNGSMSCFRNHAEGSGAISADAFSLQNYLFTAFEEENSGKNGCA 177
DB 106 PGAT-----SVSGTKNLIDISGFSL-----FSCACCPGTTGGCA 141
OY 178 IOAQTFSLSRNVSPISPARNRADLNGAICGNNLICSGNVNPLFTGNSATNGAICCTIS 237
DB 142 IOTKCTTTLKDNSSLVFHKH-----C-----STAGCAIQCKS 174
OY 238 DLNTSEKGSLSLACQOETLFAVNSAKKEKGATYAKHMYLRYNGVPSFIN-----NSAKIG 292
DB 175 SSSTAE---LKLNNKKNLVFSENSKKEGGAITYADKLTVISGPGTLFSSNVSVSHNSSPKG 231

```

```

OY 293 GAIATQ-SGGSLSLILAGCVLFPQNNQORTSDQ---LVRNAIYLEKDAIILSLAARG- 347
DB 233 GAICIKDSGECSTLANLCDITFDGKNIITTTNGGSPTYRNSIDIGSGGKFKLAKGEF 291
OY 348 DILFEDPIVOESSKESPEPSSLSQASVTSPPATASPLVLCQASRVYFSSEERLSEEEK 407
DB 292 GIFYDPDIANTGCSFEIENK-----TES-----DTTYGKIYFSEKESLDEEK 335
OY 408 T-PDNLTSOLOOPTELKSGRLVKDRAVLSAPLSQDPOALLIMEAGTSLKTSQD---L 462
DB 336 TVPANLKSIFYKQPLKIGAGSLVKDGVLEAKKITQTKGSTVVMIDGTLTQTPSSSGETI 395
OY 463 KLATSLIPHSU-----DTEKSVTHIAPNLSIQKIFLNSGDNFENYELLSEKQNN 515
DB 396 TLTNLDINIASLGGGGTAPAKLATNTASQALISIAVNLVNT-DSNTYED-PILSASK-S 452
OY 516 IPLTLTSKEQSHLHPDGNLS-----SHEGYQDMTFGSKDDEGHSLIA--NMTPKNVY 568
DB 453 FSAITATTSSTVTPPEINLKNYTPPTHGYGQNMVTVMKQSSAOEKTATLTWQOTGYS 512
OY 569 PIPEROSTIVANTLNTYSQMAVOAMINTAHGAYLEGTGSAVSNLEFAHDSSGKPI 628
DB 513 PNPBERGSLVPNTLWGAESDTRAIQNLMDISVNGADYSRGEFVSSLANFLNKSQSDTK- 570
OY 629 DWMHRSILGYLFGISTHSLDHSFCLAAQOLICKSDDSFITSTETSYIATVQAOQ---- 684
DB 571 KKFRIHSAQYALGVYAQTPSDDYVCSAFCOLFKKDKDFVSKNSSTIYAGSLTYOHSIW 630
OY 685 -----ATSLMKISAQACYNESIHELKTYRS-----FSKEGFGSMHSAVS 725
DB 631 NTWNTLLQWTLGAEPVLVMAQLTYCHASNNKTKMTNTYTPKNTYPSBKIDKQWDCQF 690
OY 726 GEVCAISIPVNSGSLFSSFSIFSKLOGSFQDGEFEE-SSGEISFSASSFRNISLPIC 784
DB 691 YEFKAKAPL-ETASLFLFDYSPFVKQLVHAHQDDEFKENNSDQGRYFESNNLTNLSMIPG 749
OY 785 IFFEKKSOKTRTYVYFLGAYIQDLKRDVEGCPVLLKNAVS--WAPAMANLDSRAVMEFL 842
DB 750 VLELFESHKDTASYMLTLAYADIRSNPDCTASLLVSTSAVWTKANNLRHAFIIOA 809
OY 843 TNQRLHRLQTLN-VSCVLRGOSHYSIDLGTTFYF 878
DB 810 GNYLATRNTLFSQFELRGSCRTYVNIIDLSKIQF 846

```

```

RESULT 17
P71132 PRELIMINARY: PRT: 847 AA.
AC P71132:
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE POMP91A.
OS Chlamydia psittaci.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RX MEDLINE=96406378; PubMed=8810511;
RA Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;
RT "Molecular cloning and characterization of the genes coding for the
RT highly immunogenic cluster of 90-kilodalton envelope proteins from the
RT Chlamydia psittaci subtype that causes abortion in sheep.";
RL Infect. Immun. 66:1317-1324(1998).

```

```
Db 700 FCHLEFARDKCFIAHNSRTYGGTLFFKHSHTLQOPNLYRLGRKFSSEAIKFPREIDL 759
Oy 691 -ISAGACVNESTHELTCKTKRSKESGSMHGVAVSGEVCASIP-IVNSGSLFSSFSIF 748
Db 760 AADVOSSESHSONRNETHTLSL-PESEGSWSNECJAGIGLDPVLSNPHPLFKTFIQ 818
Oy 749 SKLQCFSGTODGFEESGEIRSFSSAFSPFNISLPICITFEKKSOIKRTTYFLGAIQDL 808
Db 819 MVEVEMYVYQNSFEESSSDGRGFSIGRLNLSPVGAKEVQDIDGISTYDLDSGFVSDV 878
Oy 809 KADVESGPVVLKNAVSDAPMANLDSRAVYFRLLTNQRLHLQTLNANSC-----V 860
Db 879 YRNNPOSTALVMSPDMSKTRGCMNLSRQAFLLRGSN-----NYYVNSCDELFCHYAME 931
Oy 861 LRQSHSYSLDGTGTYRF 878
Db 932 LRQSSRNVDVGTCLRF 949

RESULT 13
Oy 09871 PRELIMINARY: PRT: 772 AA.
AC 09871:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
PM 3
CN PMP_3.2
OC Chlamydia pneumoniae (Chlamydia pneumoniae).
OS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shitai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and Cw1029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL: AP002545; BAA98226.1;
SQ SEQUENCE 772 AA; 82931 MW; 484FC56D35801EB CRC64;
```

Query Match 18.5%; Score 837.5; DB 2; Length 772;
Best Local Similarity 31.0%; Pred. No. 2; 7e-51;
Matches 249; Conservative 116; Mismatches 347; Indels 91; Gaps 22;

```
Oy 136 MSFCRNHAGSGGCAISADAFSLQHNLYLTFEENSSKNGCAIO-AQTESLRNVPISF 194
Db 1 MLEFSKMFSTDNCAIRAKTLSTLGTMSALFSESTSSKKGAIQTSDALITNGEVSF 60
Oy 195 ARNRADLNGAICCSNLCGAVNPLFFTGNSAT-----NGAICCIDLNTSEK 244
Db 61 SDMTSSDSGALFTTEASVVISNNAKVSFDKNKVTGASSSTTGDSGCAICA---YKTSFD 117
Oy 245 GSLSLACNELLTFASNAKEGAIAKIHVLYRNGVPVSFINNS-----AKIGAIATIOS 299
Db 118 TKYTLGNOMLLESNNITSTAGGAILYKLELSSGILTFSRNSVNGGTAHPKGAATIED 177
Oy 300 GGSLSLAGEGVALFONNSORTSDOGLVRNAYLLEKDALISLEARNG-DILFEDFIYOE 358
Db 178 SGELSLISADSGOIVFLGNVTSTTTCGNSSSIDLGTSAKMTALRSAAGRATVYDDITTG 237
Oy 359 SSS-----KSPLPSSIQASVTSPTPATSPVLIQTSANSVYFSSRLSEEEKTPD 410
Db 238 SSTVTVDVILKVNTPADSALQYT-----GNIIFTGKISLSETRADS 278
Oy 411 -NTISOLQPIELKSGRLVAKRAVLSAPLSODPOLLINEAGTSIKTSSDKLATLSI 469
Db 279 KNTSKLQDPVTLISGGLSLIKHGVTLQTOAFTQADSRLEMDVGTTLPEADTSTINNLVI 338
Oy 470 PLHSLDTEKSVTH---APNLISQKIFLNSGDENFENVELLSEKONNIPLLTIS-KE 524
```

```
Db 339 NISSIDGAKKAKIETKATSKNLISGTLTLDPTGFEYENHSLRNPOSIDILEAKSGTV 398
Oy 525 QSHLHLPOGNLSS--HFGVQGM-ITSMKDSDEGSLTL--WTKRNYPRHREQSTLYA 579
Db 399 TSTAVTPDPIKGEKFFHYGQGMPIVW---GTGASTATATFMWTGTGYIPNPERIGSLVP 455
Oy 580 NTLAMNTSDMOVQSMINTIANG--GAYLFGTMGSVSNLFVAHDSGKPIDNMHRSIG 637
Db 456 NSIMNAFIDISSLHYLMEFANEGLOGDRAP--WCAGLSNFF--HNDSTTRRGFRLSSG 511
Oy 638 YLEGISTHSLDHSFCLAAGOLLKSSDSFTTSTETSYIATVOQLATSLMKI----- 691
Db 512 YVIGLNLHTCSDKILSAACQLFGRDRDYFAKKNQCTVYGGTLYYOHNETHYISLPCKLRP 571
Oy 692 -----SAQACVNESTHELTCKTKYRSFSEEGSGSHVAVSGEVCASIPYVN 737
Db 572 CSLSYVTEIPVLFSGNLSTYHTDNDLTKYTYTPVK-GSMGNDSPALEFGGRAPICLD 630
Oy 738 GSGLFSSFSIFSKLOGFSGTODGFEESGEIRSFSSAFSPFNISLPICITFEKKSO-KTRT 796
Db 631 ESNLEFQYMPFKLOFVYVHOGEGFKGTAREFSGSLVNLALPIGIFDESDQDQT 690
Oy 797 YVYFLGAYTODLKDVEGSPVVLKNAVSDAPMANLDSRAVYFRLLTNQRLH-RLQTL 855
Db 691 YNLTIG-YTVDLVRSNPDCTTLRLISGDSMKTFGTNLARQALVLRAGNHCFNSNFARS 749
Oy 856 NVSCVLRGSHSYSLDGTTYRF 878
Db 750 QFSFELRGSSRNVDVGTCLRF 949
```

RESULT 14

```
Oy 092896 PRELIMINARY: PRT: 973 AA.
AC 092896:
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DE 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY PMP_13 PROTEIN.
GN PMP_13.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=9920606; PubMed=10192288;
RA Kalman S., Mitchell W., Marathe R., Lamme C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
```

```
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shitai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and Cw1029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL: AP002546; BAA98660.1;
SQ SEQUENCE 973 AA; 102761 MW; E02A69F611DBEF2 CRC64;
```

Query Match 18.0%; Score 816.5; DB 2; Length 973;
Best Local Similarity 26.4%; Pred. No. 1; 2e-49;
Matches 252; Conservative 158; Mismatches 404; Indels 139; Gaps 25;

```
Oy 37 NPNHYVCTFEEDCTMSLPAALCAHNS-----ODPLVYLGNSYCFVSKLHITDPKEAL 91
Db 49 DPGTCLIFSGDLYIANLNDNAISRTSSCSFNRAGALQILKGQVFSFLNIRSSADGAAT 108
```

RESULT 10
086163
ID 086163 PRELIMINARY: PRT: 914 AA.
AC 086163;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE OUTER MEMBRANE PROTEIN 5 PRECURSOR (FRAGMENT).
GN OMP5.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CDC/CWL-029/VR-1310.
RC Knudsen K., Madsen A.S., Mygind P., Christiansen G., Birkelund S.;
RT "Identification of two novel genes encoding outer membrane complex
associated surface layer proteins in Chlamydia pneumoniae."
RL Submitted (AUG-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: AJ001311; CA04671.1; -
KM Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 >914 OUTER MEMBRANE PROTEIN 5.
FT NON_TER 914 914
SQ SEQUENCE 914 AA; 95603 MM; 8BF33B4B680F5E3 CRC64;

Query Match 18.8%; Score 853.5; DB 2; Length 914;
Best Local Similarity 30.8%; Pred. NO. 2.5e-52;
Matches 278; Conservative 137; Mismatches 382; Indels 107; Gaps 30;

QY 12 CAALISSTAVLFGDPL--GETALLTKNPNHYC-----TFEDCTMESL--FPALCAHAS 63
DB 20 CSTVAAATRENIGPSDSDGCTNTCTYTPKNTTGTIDYTLGCDITLQNGDSALTKGCF 79
QY 64 OD-DPLVVLGNSYCMFVSKLITDPKALFEKGDLSIONFRPLSTDCSKE--SSPS 119
DB 80 SOTTELSLPAGKYS--LSFLNKSAG-----AALSVTDKNLSLGFSSLTFLAPS 132
QY 120 II-----HKNQGLSLRNGSMSCRNHAECSGCAISADAFSLQHNVIPTAFENS 170
DB 133 SVTTTPSKGAYKCGDITLFDNNGTILFKODYCEENGCAISTKNLSLKNSTGSISEFECN 192
QY 171 SKG---NGGAIDAO--TFLSRNVSPISFARNRADLNGAI--CCSLICSGNVPLEFFGN 225
DB 193 SSATCKKGGALCATCTVDTTNTATFLFSNNIAEAGCALINSTGCTTGNTS-LVFSN 251
QY 226 SAT---NGAICICISDNTSEKSLSLACNOETLFAASNAKEKGATYAKHMYLRY--N 279
DB 252 SVTATAGNGAL-----SGDAVTISNGSVTFSGNQAANAGAIYAKKTLTASGG 303
QY 280 GPVSTINN-----SAKIGALAIQSGSLILAGEGSLVLFQNN--QRTSDGLVNAIYL 333
DB 304 GGISFSNNIVOGTTAGNGAISILAGCSCSLAEAGDITFNCNAIVATTPTTKRNSIDI 363
QY 334 EKDALISSEARNG-DILFFPIVQESSKESPLPSLOASTPTPTATASPLVIQTSAN 392
DB 364 GSTAATITNRAISGHSIFFYDITANTADSDITLNLKADAGN-----STDSY 412
QY 393 RSVIFSSERLSEEE-KTDPNLTSLQLOPIELKSGRLVLDRAVLISAPLSODPOLLINE 451
DB 413 GSIVFSGEKLSEDEAKVADNLSTLKQPVTLTLAGNLVLRGTLDLTQGTOTAGSSVIMD 472
QY 452 AGTSUKTSS-DLKLTLTSLPLSHSLDTEKSVTHAP-----NLSTQKIFLSNGGENTYE 504
DB 473 AGTTLKASTEEVTLTGLSTPVDLSLEGKKVYLAASAASKNNALSPILLDNOG--NAYE 530
QY 505 NVELLSKEON-----NIPLLTSLKEOSHLPLPDGNSSHFGYGGDTFSWKD-----SDEGH 556
DB 531 NHD-LGKTODPSFVQLSALGTATTTTDDPAVPTVATPTHTKYGCTGKMTAVDQDTASTPKTK 589

QY 557 SLIANWTPKNTVPPEROSTLVANLTWNTYSDMOAVOSMINTIAHGAYLFTGWSAVSN 616
DB 590 TATLATNTGYLPNPERGCLPVNSLMGSEFSDIAQGVIERSAHLTLCSDGFGAAAGVAN 649
QY 617 LFVADSSGKPIIDNMHRSGLYLFQISTHSLDHSFLCAAGQLKSSDSFTSTETTSY 676
DB 650 -FLDKDKKGEK-RKYRHKSGGAYIGAQAOTCSENLSFAFCOLFGSPKDFLVANHTDTY 707
QY 677 IATVQAOLATL-----MKISAQCYMESIHELKTKRYSKSGFG 717
DB 708 AGATYQIHTSCSGFIGCLDLKPSMSHKPLVLEGQLAYSHVSDNLTXTATY-PEVK 766
QY 718 SWHSAVAGVECAPIIVSNCGSLFSSFSIFSKLOGFSGTODCFEESGSEIRSSASSFR 777
DB 767 SWGNNAFNMMLGASHSYPERVLCEDTVAPIKTLNLTYYIRNDSFSEKTEBRSFDDENLF 826
QY 778 NISLPIGITEFKKSOKTRTYTFGLGATYQDLKRDYESSPVVLLKNAVSWDAPMANLDSRA 837
DB 827 NLSLPIGYKFEKFSDCNDFSYDLTSLYVPLDIRNDPKCTTALVISGASWETVANNLARO 886
QY 838 YMR 841
DB 887 LQVR 890

RESULT 11
086164
ID 086164 PRELIMINARY: PRT: 928 AA.
AC 086164;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE OUTER MEMBRANE PROTEIN 4 PRECURSOR.
GN OMP4 OR PWP.11.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CDC/CWL-029/VR-1310;
RC Knudsen K., Madsen A.S., Mygind P., Christiansen G., Birkelund S.;
RT "Identification of two novel genes encoding outer membrane complex
associated surface layer proteins in Chlamydia pneumoniae."
RL Submitted (AUG-1997) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=VR1310;
RC Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
RA Christiansen G., Birkelund S.;
RT "Chlamydia proteins containing the GGA1-repeat belong to a subfamily
of autoantigenic pathogenicity factors."
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206066; PubMed=10192388;
RA Kaiman S., Mitchell W., Marathe R., Lamme C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."
RL Nat. Genet. 21:385-389(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CWL029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL: AJ001311; CA04672.1; -
DR EMBL: AJ133034; CAB37072.1; -
DR EMBL: AE001628; AAD18593.1; -
DR EMBL: AP002546; BAA98658.1; -

121
RN SEQUENCE FROM N.A.
RP STRAIN-AR39:
RC MEDLINE-20150255: PubMed-10684935:
RA Read T.D., Brinham R.C., Shen C., Gill S.R., Heidelberg J.F.,
White O., Hickey E.K., Peterson J., Umayam L.A., Ulterback T.,
Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J.,
McClarty G., Salzberg S.L., Eisen J., Fraser C.M.:
RT "Genome sequences of Chlamydia trachomatis Morn and Chlamydia
pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL: AP002546; BAA98652.1; -
DR EMBL: AE002193; AAF38166.1; -
DR TIGR: CP0309; -
SO SEQUENCE 1276 AA; 132126 MW; C97FA40662C88E0B CRC64;

Query Match: 19.7%; Score 894.5; DB 2; Length 1276;
Best Local Similarity 30.6%; Pred. No. 5e-55;
Matches 279; Conservative 144; Mismatches 364; Indels 125; Gaps 29;

QY 68 LYVLNVCWCPFSKL-----HTDPKEALFKKGLDSTONFFLSFTDSSKSSPSSTIH 122
DB 389 LYSKGNSSLTGTNTLLFSGCNKATGPNSSANQEG---CGAILSLFLESAS-----VS 437
QY 123 OKNGOLSLRNNGSMFCRHHAGSGAISADAFSLQHNLYLTFAPFENSCKGNGAIOAQT 182
DB 438 TKKG-LWIEDNENVSLSGNTATVSGGAIYATKCALHGNTTLT-FGQNTAETAGAIYET 495
QY 183 -FSLSRNVSPISFARNRADLNGAICCSNLICSGNVNPLFTGTSATN----- 229
DB 496 EDFTLTGSTGTFTSTNKTAKTAGALHTKGTSTFN-KALVFSGNSATATATTDTQEGC 554
QY 230 GCAIICC-ISDLNTESEKGSLSLACNDELTFANSKE----- 264
DB 555 GCALICNISESDIATK-SLTLTENESLSFINNTAKRSGGCIYAPKCVISGESINFDGNT 613
QY 265 ---KGAIAIKHMYLRYNGVPVSPFINNSAKIGAIAGSGLSILAGESSVLFQNNQORT 321
DB 614 AETSGAIYKSKULSTANGPVSFTNNSGKGGAITYADSGELSLADIDITFSGN--RA 671
QY 322 SDQGLVRNAILKENDALLSLEARNG-DLFFDPDIVOESSS-----KESPLSSLOASVT 375
DB 672 TEGTSTPNSIHLAGAKAKITKLAAPAGHTIYFYDPTMEAPAGCTIEELVINPVKATVP 731
QY 376 SPT-----PATASPLYIQTSA--RSVIFSSERL-SEEKTPDNLTSOLOPTELSGRL 427
DB 732 PPOPKNGPIASVPVAPAPANPTGTIVFSSGKLPSODASIPANTTTILNOKINLAGCNV 791
QY 428 VLKDRAVIASPLSODPOLLIMEAGTSLKTS-----DKLATLSIPLSLDTEKSVT 481
DB 792 VLKEGATLOVYSFTQOPDSTYFMDAGTLETTTNNNGSIDLKLSVNLADDKRMIT 851
QY 482 I---HAPNLSTQIKIFLSNSGDENFYENVELLSKEQNNIPLLTJLSEKSHLPLDGN--- 534
DB 852 IAVNSTSGGLKISGDLKFHNNEGSFYDNPGT--KANLNLPLFDLSTSGTVLNDPFNPLP 909
QY 535 ---LSHFGYOGDMTFMSKDSDEGH-SLIANTPKNYVPHPEROSTLVANLTMNTYSDMQ 590
DB 910 SSMAAPDVGYSQSWTLVPRVAGAGVTLVAEWAQALGYTPKRLATLPLPNSLMNAVYNIH 969
QY 591 AVQSMINTIAGGAYLFTWGSVAVSNLFYAHDSGKPLIDNMHRSLSGLYFGISTSLDDH 650
DB 970 SIOQGIATAMSDAPSPHPIGWIGIGNAF--HODKOKENAGRLIRGIIYVGSMTTPOEY 1027
QY 651 SFCLAGOLKSSDSFTTSTETTSYIATVQAQ-----LATSLMK----- 690
DB 1028 TFAVAFSOLFSGSKDYVVDISKQYVAGSLCAOSSYVILPLHSRLRHVLSKVLPLPGET 1087
QY 691 ---ISAQACYNSTIHELTKYRSFSEKGFSGHSAVAGEVCASIPYISNGSGLFSSSSI 747
DB 1088 PVLHGQVSYGRNHHNMTTKLAN-NTQGRKSDMDSHSFAVEGSGSLPVDLNTRYL--TSYSP 1145

QY 748 FSKLOGFSGTODGFEESSEGEIRSFSSASFRNLSLIGITFEKKSKOTRYYFLCATYOD 807
DB 1146 YVKILOVSYNOKGFOEVAADPRIFDASHLVNYSIPMGITFEKIESAKKPSALLTLGLYVD 1205
QY 808 LKRDESGPVLLKNAVSDAPMANLDSRAYMFRLTNO-RALHRLQTLINVCVLRGQSH 866
DB 1206 AVRQ-HPHCLTSLTGTSTMSRTATLNSQAFPAEASGHLKHLGLDCAAGSGCELRSSR 1264
QY 867 STSLDLGTTTYR 878
DB 1265 SYNANGTRYSP 1276

RESULT 8
ID Q92899 PRELIMINARY; PRT: 1407 AA.
AC Q92899;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G/I FAMILY.
GN PMP-6.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CWL029;
RX MEDLINE-99206606; PubMed-10192388;
RA Kallman S., Mitchell W., Marathe R., Lammell C., Fan J., Hyman R.W.,
R. Olinger L., Grimwood J., Davis R.W., Stephens R.S.,
"Comparative genomes of Chlamydia pneumoniae and C. trachomatis."
RL Nat. Genet. 21:385-389(1999).
DR EMBL: AE01627; AAD18588.1; -
SO SEQUENCE 1407 AA; 144919 MW; 9CF54AA4614D54C7 CRC64;

Query Match: 19.7%; Score 894; DB 2; Length 1407;
Best Local Similarity 31.1%; Pred. No. 6.3e-55;
Matches 270; Conservative 140; Mismatches 343; Indels 116; Gaps 27;

QY 106 LSFDCSSKESPSLIHQKNOLSLRNNGSMFCRHHAGSGGAIADAFSLQHNLYLFTA 165
DB 559 LSFLESAS-----VSTRKG-LWIEDNENVSLSGNTATVSGGAIYATKCALHGNTTLT- 609
QY 166 FEENSSKNGAIOAQT--FSLSRNVSPISFARNRADLNGAICCSNLICSGNVNPLFT 223
DB 610 FDGNTAETAGAIYETEDFTLTGTGTFTSTNKTAKTAGALHTKGTSTFN-KALVFS 668
QY 224 GNSAN-----GCAICC-ISDLNTESEKGSLSLACNDELTFANSKE----- 264
DB 669 GNSATATATTDTQEGCGGAILCNISESDIATK-SLTLTENESLSFINNTAKRSGGIYA 727
QY 265 -----KGAIAIKHMYLRYNGVPVSPFINNSAKIGAIAGSGLS 304
DB 728 PKCVISGESINFDQNTAETSGAIYKSKULSTANGPVSFTNNSGKGGAITYADSGLS 787
QY 305 ILAGGSVLPFONNSORTSDQGLVRNAILKENDALLSLEARNG-DLFFDPDIVOESSS-- 361
DB 788 LEALIDITFSGN--RATEGTSTPNSIHLAGAKITKLAAPAGHTIYFYDPTMEAPASG 845
QY 362 ---KESPLSSLOASVTPT---PATASPLYIQTSA--RSVIFSSERL-SEEKTPD 410
DB 846 GTIEELVINPVKATVPPOPKNGPIASVPVAPAPANPTGTIVFSSGKLPSODASIPA 905
QY 411 NLTSOLOPTELSGRLVLKDRAVIASPLSODPOLLIMEAGTSLKTS-----DKL 464
DB 906 NNTTLNKNINLAGSNVYLKREGATLOVYSFTQOPDSTYFMDAGTLETTTNNNGSIDL 965
QY 465 ATLSIPLSLDTEKSVT---HAPNLSTQIKIFLSNSGDENFYENVELLSKEQNNIPLLT 520
DB 966 KNLVNLDALDQGRMITIAVNSTSGGLKISGDLKFHNNEGSFYDNPGT--KANLNLPLFD 1023

OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-J138:
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CWL029 from USA."
 RL Nucleic Acids Res. 28:2311-2314(2000).
 RP SEQUENCE FROM N.A.
 RC STRAIN-AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,
 RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Green B.,
 RA Bowman C., Dodson R., Gwinn M., Nelson W., Deboy J., Kolonay J.,
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
 RT pneumoniae AR39."
 RL Nucleic Acids Res. 28:1397-1406(2000).
 DR EMBL: AP002546; BAA98654.1;
 DR EMBL: AE002193; AAF38104.1;
 DR TIGR: CP0307;
 SQ SEQUENCE 930 AA: 97639 MW: 468896761391C09 CRC64;

Query Match 20.3% Score 919.5; DB 2; Length 930;
 Best Local Similarity 31.0%; Pred. No. 5.2e-57;
 Matches 260; Conservative 150; Mismatches 343; Indels 87; Gaps 26;

OY 98 LSIQNFRLSF-----TDCSSKESPSTIIHOKNQLSLRNGNSMFCR---NHAEGSGGA 149
 DB 119 LFTFGFSLNSTIAAPGTIVAGSKSTLS---SAGALNTDNGTILFSGQNSNEANNNGA 174
 OY 150 ISADAFSLQHHYLTFAFEENSSKNGCAI-OAQTFSLRNVPISFARNRADLNGCAICC 208
 DB 175 ITAATLTSIGNTSSITFTFSNSAKKLGAIYSSAASISGNTGQLVFNKKEGTGCGAL- 232
 OY 209 SNLJCSGVN---NPLFTGNSATN---GGAICISDLNTSEKGSLSLACQETLFAENS 261
 DB 233 -GFEASSITONSSLFESGNTATDAAGKGAICY--EKTEPTLTITSGKKSILTFAPENS 288
 OY 262 AKERGAIAIAKHYLVRYNGPVSFINN-----SAKIGAIAGOSGSLSLIAGEGSLVFON 316
 DB 289 SVTGGGAIACAGLDLSAAGPTLFSNNRCGNTAAKGGAIAIADSGSLSLANOGDITFLG 348
 OY 317 NS-ORTSDQGLVRNAIYLEKADILSLSEARNG-DILFDPPIVOESSKESPLPSLOASV 374
 DB 349 NTLTSTSAFTSTRAIYLGSSAKITNLRAAGOSIYFDPPIASMTGTASDYL----- 400
 OY 375 TSPPTATSPPLVIQTSAARSVIFESSERLS-EBEKTPDNLTSOLOQPIILKAGRLVLDRA 433
 DB 401 TTNOPDSNPL-----DVSGTIVFSGEKLSADEKAADAFSTILKOPILASGLTALKGNV 456
 OY 434 VLSAPSLSDQPOALIMEAGTSLKTSSTP-LKLATLSILPLHSLDTEKSVTIAP-----N 486
 DB 457 ELDVNGFQTGCTSLLMQPGTKLKADTAISLTILKYVDLSLDEGKKSISTETAGANKTIT 516
 OY 487 LSIQKIFLNSGDENFENVELLSKEDONIPPLTLLSKEQSHLHPDGNLSS-----HF 539
 DB 517 LTPSLVYDQSSG--NFYESHTINQAFOTPLVFTAAFTAAISDIYI-DALITSPVOTPEPHY 573
 OY 540 GYGDWTFESMNDSEGHLLIANMTPKNVVPHPEROSTLVANTLWNTSDMAYOSMINTI 599
 DB 574 GYCGHWETATADTSTAKSGTMTWTTCINPPEKASAVPDSLWASFDITDLOQIMTSQ 633
 OY 600 AHGCAVLTGTCGSAVSNLFVAHDSGKPIDMWHRSILQYLEGISTHSLDHSFLCAACQL 659
 DB 634 ANSIYQORGLTASGTANFEH-KDKSGTN-QAFRRKSYGIYVGSASDEPSENIFFVAFCOL 691
 OY 660 LGRSSDSFTISTETTSYATATVQAO-----LATSLMK-----ISAQACYNES 700

DB 692 FCKDKDLFTVENTSHNYLASLYLQHRAFLGGLPMPSFGSITDMLKDPIILLNAQLSYST 751
 OY 701 IHELTXYRFSRKEGFGWHSVAVSGEVCASPT-VSNGSGLFSSFSIFSLQGFSGTOD 759
 DB 752 KNDMDTRTSY-PEAOGSGWNTNSGALLEGSLATLPLEAEFFQGFPPFLFOAVYSRQ 810
 OY 760 GFEESGEIIRFSASFNPISLPIGIFTEKKSOKRTTYFLGAVIODLKRDEVESGPVYL 819
 DB 811 NFKESGAEARAFDGDGLVNCSTIPVGRLEKISEDEKNNEFISLAVIGVYRKNPRTSL 870
 OY 820 LKNVSMDPANIDSRVYMERLTNQRALH-RLQTLNVSCVLRGQSHSYSLDGLTYRF 878
 DB 871 MVSGASWTSLSCKNLARQAFASAGSHLTLSPHVELSGAAVEILNGSAHYIVDGLKRSF 930

RESULT 5
 ID 092393 PRELIMINARY; PRT: 930 AA.
 AC 092393;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
 DE OUTER MEMBRANE PROTEIN 11 PRECURSOR.
 GN OMP11 OR PMP-8.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CWL029;
 RX MEDLINE=99206606; PubMed=10192388;
 RA Kaiman S., Mitchell J.W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."
 RL Nat. Genet. 21:385-389(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VIR310;
 RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
 RA Christensen G., Birke Lund S.;
 RT "Chlamydia proteins containing the GGA1-repeat belong to a subfamily
 RT of autotransporting pathogenicity factors."
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ133034; CAB37068.1;
 KW Signal.
 FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 930 OUTER MEMBRANE PROTEIN 11.
 SQ SEQUENCE 930 AA: 97669 MW: 46A9B5E3BB913C4C CRC64;

Query Match 20.2% Score 915.5; DB 2; Length 930;
 Best Local Similarity 30.8%; Pred. No. 1e-56;
 Matches 259; Conservative 150; Mismatches 344; Indels 87; Gaps 26;
 OY 98 LSIQNFRLSF-----TDCSSKESPSTIIHOKNQLSLRNGNSMFCR---NHAEGSGGA 149
 DB 119 LFTFGFSLNSTIAAPGTIVAGSKSTLS---SAGALNTDNGTILFSGQNSNEANNNGA 174
 OY 150 ISADAFSLQHHYLTFAFEENSSKNGCAI-OAQTFSLRNVPISFARNRADLNGCAICC 208
 DB 175 ITTATLTSIGNTSSITFTFSNSAKKLGAIYSSAASISGNTGQLVFNKKEGTGCGAL- 232
 OY 209 SNLJCSGVN---NPLFTGNSATN---GGAICISDLNTSEKGSLSLACQETLFAENS 261
 DB 233 -GFEASSITONSSLFESGNTATDAAGKGAICY--EKTEPTLTITSGKKSILTFAPENS 288
 OY 262 AKERGAIAIAKHYLVRYNGPVSFINN-----SAKIGAIAGOSGSLSLIAGEGSLVFON 316
 DB 289 SVTGGGAIACAGLDLSAAGPTLFSNNRCGNTAAKGGAIAIADSGSLSLANOGDITFLG 348
 OY 317 NS-ORTSDQGLVRNAIYLEKADILSLSEARNG-DILFDPPIVOESSKESPLPSLOASV 374


```

QY 181 QTFSLSRNVSPISFARNRADLNGAICCSNLICSGNVNPLFTFGTSATNGAICCIDSLN 240
DB 181 QTFSLSRNVSPISFARNRADLNGAICCSNLICSGNVNPLFTFGTSATNGAICCIDSLN 240
QY 241 TSEKGSLSLACQOETLPAFNSNAKEKGAIYAAHMYLRYNGPVSFINNSAKIGCAIAIQSG 300
DB 241 TSEKGSLSLACQOETLPAFNSNAKEKGAIYAAHMYLRYNGPVSFINNSAKIGCAIAIQSG 300
QY 301 GSLSLIAGEGSVLPONNSORTSDQGLVRYNAIYLEKDAIISLSEARNGDILFFDPVIOESS 360
DB 301 GSLSLIAGEGSVLPONNSORTSDQGLVRYNAIYLEKDAIISLSEARNGDILFFDPVIOESS 360
QY 361 SKESPLPSLQASVTSPPATASPLVIQTSANRSVIFSSERLSEEEKTPDNLTSOLOQPI 420
DB 361 SKESPLPSLQASVTSPPATASPLVIQTSANRSVIFSSERLSEEEKTPDNLTSOLOQPI 420
QY 421 ELKSGRLVKDRAVLSASLSODPQALLIMEAGTSLKTSDDLKATLTSIPLHSLDTEKSV 480
DB 421 ELKSGRLVKDRAVLSASLSODPQALLIMEAGTSLKTSDDLKATLTSIPLHSLDTEKSV 480
QY 481 TIHAPNLSIOKIFLSNCGDENFEYENVELLSKEONNIPLLTLSKEQSHLHPDGNLSHFG 540
DB 481 TIHAPNLSIOKIFLSNCGDENFEYENVELLSKEONNIPLLTLSKEQSHLHPDGNLSHFG 540
QY 541 YOGDWTFSWKSDSGHSLIANWTPKRYVPHPEROSTLVANTLMTYSDMAVQSMINTIA 600
DB 541 YOGDWTFSWKSDSGHSLIANWTPKRYVPHPEROSTLVANTLMTYSDMAVQSMINTIA 600
QY 601 HGGAVLFTGMSAVSNLRYAHDSSGKPIDNMHRSGLGFLGISTHSLDDHSCLAAGQL 660
DB 601 HGGAVLFTGMSAVSNLRYAHDSSGKPIDNMHRSGLGFLGISTHSLDDHSCLAAGQL 660
QY 661 GKSSDSFITTSTETSYIATVOAQLATSLMKISAQACYNESIHELKTKRFSKEGFGSMH 720
DB 661 GKSSDSFITTSTETSYIATVOAQLATSLMKISAQACYNESIHELKTKRFSKEGFGSMH 720
QY 721 SVAVSGEVCASIPYVNSGGLFSSFSIFSKLQFGSGTODGFEESGSEIRSFASAFRNI 780
DB 721 SVAVSGEVCASIPYVNSGGLFSSFSIFSKLQFGSGTODGFEESGSEIRSFASAFRNI 780
QY 781 LPIGTFFKKSKOKRTYRYFLGAYIYODLKRVEGSPVLLKANVSWDPMANLDSRAFMF 840
DB 781 LPIGTFFKKSKOKRTYRYFLGAYIYODLKRVEGSPVLLKANVSWDPMANLDSRAFMF 840
QY 841 RLTORALHRLQTLNLVSCVLRCQSHSYSLDLGTYRF 878
DB 841 RLTORALHRLQTLNLVSCVLRCQSHSYSLDLGTYRF 878

```

```

RESULT 2
Q9P141 PRELIMINARY: PRT: 867 AA.
AC Q9P141:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE POLYMORPHIC MEMBRANE PROTEIN G FAMILY.
GN TC0267.
OS Chlamydia muridarum.
OC Bacteria: Chlamydiales: Chlamydiaceae: Chlamydia.
OX NCBI_Taxid=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / NIGG:
RX MEDLINE=20150255: Pubmed=10684935:
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Ullterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gilm M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.:
RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RA Read T.D., Brunham R., Shen C., Gill S.R., Heidelberg J.F., White O.,
RA Hickey E.K., Peterson J., Umayam L.A., Ullterback T., Berry K.,
RA Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,
RA Dodson R., Gilm M., Nelson W., Deboy R., Kolonay J., McClarty G.,
RA Salzberg S.L., Eisen J., Fraser C.M.:
RA Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RL EMBL: AE002294; AAF39136.1;
DR TIGR: TC0267;
SQ SEQUENCE 867 AA: 9501 MW: 557994185A9E5652 CRC64;

Query Match 78.2%; Score 3543.5; DB 2; Length 867;
Best Local Similarity 77.8%; Pred. No. 5.8e-244;
Matches 684; Conservative 76; Mismatches 106; Indels 13; Gaps 2;

QY 1 MRPDHMFCCCAATLSTAVLFEGDPLGETALTLPKNHVCCTFEEDCTMESLFPALCA 60
DB 1 MRPDHNVLCCLATILSTPTALLFGODALDKSALITKPNNSIVCTFLEDCTMENSPALLS 60
QY 61 HASODPLVUNGSYCMFVSKLHITDPKEALFEKGDLSIONFRFLSTDCSSK-ESSPS 119
DB 61 HARODDPLYIIGNTHMVEVSNLHPSTNEERFLKEGDLSDPFRFLSTDCSSSTEDSPS 120
QY 120 ITHKNGQLSRNNGSMFCRNHAEKGSADAFSLQHHYLTFAPEENSKNGCAIQ 179
DB 121 ILYHNGQLFLRNNGSMFYNHSEKGSALSTDALFLQHHYLTFAPEENSAKNGCAIQ 180
QY 180 AOTFSLSRNVSPISFARNRADLNGAICCSNLICSGNVNPLFTFGTSATNGAICCIDSL 239
DB 181 AOTLSTSRNVSSLSFARNRADLNGAICCCNLICSGNVNPLFTFGTSATNGAICCIDSL 240
QY 240 NTEKGSLSLACQOETLPAFNSNAKEKGAIYAAHMYLRYNGPVSFINNSAKIGCAIAIQ 299
DB 241 NLSEKGCISLAVNOETLFPNSNAKEKGAIYAAHMYLRYNGPVSFINNSAKIGCAIAIQ 300
QY 300 GGSLSIAGEGSVLPONNSORTSDQGLVRYNAIYLEKDAIISLSEARNGDILFFDPVIOEV 359
DB 301 GGSLSIAGEGSVLPONNSCHFSODGTVRYNAIYLEKDAIISLSEARNGDILFFDPVIOEV 360
QY 360 SKESPLPSLQASVTSPPATASPLVIQTSANRSVIFSSERLSEEEKTPDNLTSOLOQ 419
DB 361 VSPFSTTSAL-----TFLRIQTNRAVIFSSNLSEEEKTEENLSKIQQP 408
QY 420 IELKSGRLVKDRAVLSASLSODPQALLIMEAGTSLKTSDDLKATLTSIPLHSLDTEK 479
DB 409 IELSGCLVLEKDRVILSAPSLSQAPQALLVMDVGTSLTSSDLKLTLSIPLHSDTEK 468
QY 480 VTIHAPNLSIOKIFLSNCGDENFEYENVELLSKEONNIPLLTLSKEQSHLHPDGNLSH 539
DB 469 VSIOSPLTSLQKIFLSNCGDENFEYENVELLSKDDKDIPLSLPGLPHPLPDNLSSHF 528
QY 540 GYGDWTFSMKSDSGHSLIANWTPKRYVPHPEROSTLVANTLMTYSDMAVQSMINTI 599
DB 529 GYGDWTFSMWTSQDRETLVANTYANRYPHPERQSLVANTLMTYSDMAVQSMINTI 588
QY 600 AHGCAVLFGTWGSVSNLFAHDSGKPIDNMHRSGLYFLGISTHSLDDHSCLAAGQL 659
DB 589 AOGGAVLFTWGSVSNLFLYSHGNSGKSTDNMKHRSGLYFLGISTHSLDDHSCLAAGQL 648
QY 660 LGKSSDSFITTSTETSYIATVOAQLATSLMKISAQACYNESIHELKTKRFSKEGFGSM 719
DB 649 FGKSSDSFVTSADTFTTIAAIQIOTIATSLIKISAQACYNESIHELKTKRFSKEGFG 708
QY 720 HSAVSGEVCASIPYVNSGGLFSSFSIFSKLQFGSGTODGFEESGSEIRSFASAFRNI 779
DB 709 HSAVSGEVCASIPYVNSGGLFSSFSIFSKLQFGSGTODGFEESGSEIRSFASAFRNI 768
QY 780 SLPIGTFFKKSKOKRTYRYFLGAYIYODLKRVEGSPVLLKANVSWDPMANLDSRAFM 839
DB 769 SLPIGTFFKKSKOKRTYRYFLGAYIYODLKRVEGSPVLLKANVSWDPMANLDSRAFM 828

```

Mon May 7 09:06:43 2001

us-09-677-752-4.rst

Page 24

us

```

149 lailleserlaaspheleuGlnHisAsnTyrLeuphetrAla 165
      :::::
244 .....CTTCTGTATATAAGTTCTATAGCAAGCG 276
      :::::
166 phegluGlnAsnSerSerlysglysnGlyAlaIleGlnAlaInh 182
      :::::
277 GGATCAACAACACTCTCCAGGAGAGGTGG.....CGCGG 314
      :::::
182 rheserleuSerArgAsnValSerProIleSerPheAlaArgAsnArg 199
      :::::
315 CTATACACTATACCGAGGGGACACCC..... 342
      :::::
199 laaspleuSnGlyAlaIleCysCysSerAsnLeuIleCysSergly 215
      :::::
343 .....TGTCATAAAGTGTCT..... 360
      :::::
216 AsnValAsnProleuphetrGlyAsnSerAlaThrAsnGlyAla 232
      :::::
361 .....GTGCGCTCAGAGAGCGC 377
      :::::
232 alle.....CysCysIleSerAspleu...AsnThrSerglySglys 246
      :::::
378 TCTCTGAAACTGTGCTGTGAGAGCTGTGACACACCGAGAGGTGGAG 427
      :::::
246 erleuSerleuAlaCysAsnGlnGluThrleupheAlaSerAsnSerAla 262
      :::::
428 GCGAGTCTCTGTGTGTTCACCCCAACAGAGTGT.....GACAGC 471
      :::::
263 LysGluSgLyGlyAlaIleTyrAlaIleHisMetValLeuArgTyrAs 279
      :::::
472 GAAAGAGAGAGGTGTGTGCGGTGCGCACACAAAGTCTTATCCATGTG 521
      :::::
279 nGlyProValSerPheIleAsnSerAlaIleGlyAlaIleAla 296
      :::::
522 TGGGCCCCCGAGAGAGTATATCTGTGAAGAGTATCTCTGTGCGGGG 571
      :::::
296 laIleGlnSerGlySergleuSerIleleuAlaGlyGlySerVal 312
      :::::
572 GCACAATCTATATAAGCTGTGTGCGAGCGCCCAAGACACTCTCT 621
      :::::
313 leupheGlnAsnSerGlnArgThrSerAspGlnGlyLeuValArgAs 329
      :::::
622 CTCTTTACATTTCTGTCGAGAGACCGTCTAGGGGTT..... 663
      :::::
329 naIaIleTyrleuGluLysAspAlaIleleuSerSerleuGluAlaArg 346
      :::::
664 .....GAGACATCATTTTGGCGCCCAACCAAGCGCGC 697
      :::::
346 snGlyAspIleleupheAspProIleValGlnIleSerSerSerLys 362
      :::::
698 GC.....TCACCAAAAGTATCGCGCGGACACAGGA 729
      :::::
363 GluSerProleuproSerSerleu.....GlnAlaSerValThrSerPr 377
      :::::
730 GGGGCGCCCCCGCAAGCGTCTGTGAGGAGGCTCATATACGCGCC 779
      :::::
377 oThrProAlaThrAlaSerProleuValIleGlnThrSerAlaSnArgS 394
      :::::
780 ACTGATGATGTGTCGCCCAACATGCGCGCCCAACAAACAGAGAGAGG 829
      :::::
394 erValIlePhe.....SerSergluArgleuSergluGluLys 407
      :::::
830 TGTTCACACCGCGAGATTAGAGAGAGAGATTATCTTCAACAAGCT 879
      :::::
408 ThrProAsp.....AsnleuThrSerglnleuGlnInPro.. 419
      :::::
880 TGTCTTCTCCCGCGAGAGAAACCGAAGACTTGAAGAGAGACACACC 929
      :::::
420 .....IleGluLeuLysSergly... 425
      :::::
930 TCTAGTATGACTGGCGCCCTCTCAACACCCCTATTTCGGGCGCC 979

```

```

426 .....ArgleuValleuLys 430
980 ACTCGGGTGCCACCACAGCTTTGGAAGACACAGCTTTTAAGAAA 1029
      :::::
431 AsparGAlaValleuSerAlaProSerleuSerGlnAspProGlnAla 447
      :::::
1030 GAAAGAGACATCTTCCGGGGGCCCAACAGTGAAGAGACCCACCAAT 1079
      :::::
447 uleuIleMetGluAlaGlyThrSerleuLysThrSerSerAspLeuLys 464
      :::::
1080 CTCTTGGGGGTTCAGAAACATTTCTATGAAGAGCGCCCAATCAAGA 1129
      :::::
464 euAlaThrleuSerIleProleuHisSerleuAspThrGluLysSerVal 480
      :::::
1130 CCGGACCAATA.....CTATGACACACAGCGGACACC 1161
      :::::
481 ThrIleHisAlaProAsnleuSerIleGlnLysIlePheleuSerAsn 497
      :::::
1162 TTACTCCACAGCCCTCAC..... 1179
      :::::
497 rGlyAspGluAsnPheTyrGluAsnValGluLeuLeuSerLysGluGln 514
      :::::
1180 .....ACCACAGACAGCACTTCTGTGAGAGC 1207
      :::::
514 snAsnIleProleuLeuThrleuSerLysGluGlnSerHisleuHisle 530
      :::::
1208 AGAAACACCCCTCTCTAGAGGAGAGCGTTAGAGAGACACACACCT 1257
      :::::
530 uProAspGlyAsnleuSerSerHisPheGlyTyrGlnIleAspThrPr 547
      :::::
1258 AGTGAAGGGAAGACCAATGTCATTCATCCGCCGCCCATTTGACAC 1307
      :::::
547 heserTrpLysAspSerAspGlyHisSer..... 557
      :::::
1308 ACTGTGAGGAGATATATCCCGCACACCGCACAGTACGCGCTTCGG 1357
      :::::
558 .....leuIleAlaAsnTrpThrProLysAsnTyrValPr 569
      :::::
1358 YCGAACACACCCCTACTTGTGGCACA.....CACAAACAGTCC 1398
      :::::
569 oHisProGluArgGlnSerThrleuValAlaAsnThr 581
      :::::
1399 ACCCGCGAAGACCTCTCTGTGTGAGTGGACACAC 1435
      :::::
seq_name: gb_est71:BE215233
seq_documentation_block:
LOCUS BE215233 1831 bp mRNA EST 25-JUL-2000
DEFINITION HV_CEB0006D19f Hordeum vulgare seedling green leaf EST library
HVCN0005 (Erysiphe infected & control) Hordeum vulgare cDNA clone
HV_CEB0006D19f, mRNA sequence.
ACCESSION BE215233
VERSION BE215233.1 GI:8902845
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare
Fukuroyoka; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae;
Hordeum.
REFERENCE 1 (bases 1 to 1831)
Wing, R., Close, T.J., Kleinholz, A., Wise, R., Begum, D., Frisch, D., Yu
X., Anderson, H., Dale, J., Henry, D., Kernode, S., Palmer, M., Rambo
T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D.W., Main, D. and
Wood, T.
Development of a genetically and physically anchored EST resource
for barley genomics
Unpublished (2000)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293

```

FEATURES
source

Location/Qualifiers
1. 785
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"

/clone_lib="C. reinhardtii CC-1690, stress condition I, normalized, lambda Zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2: XhoI; This library, constructed by John Davies and Jeffrey Mcdermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min, 1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with Exassist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 352 a 1 c 411 g 17 t 4 others
ORIGIN

alignment_scores:
Quality: 99.00 Length: 221
Ratio: 0.853 Gaps: 8
Percent Similarity: 52.489 Percent Identity: 28.507

alignment_block:

US-09-677-752-4 x BF86549/rev ..

Align seg 1/1 to reverse of: BF86549 from: 1 to: 785

```

349 Tleleuphephasprolilevalingluser..... 359
729 CTCCTCTCTCTCCCTCTCTACTCTCTCTCTCTCTCTCTCTCTC 660
360 .....SerSerLysGluSerProLeuProSerSerLeuGlnAlas 373
679 TCTTATCTCTCTCTCTCTCCCTCCCTCCCTCTCTCTCTCTCTCT 630
373 erValThrSerProThrProAlaThrAlaSerProLeuValIleGlnThr 369
629 CCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 580
390 SerAlaAsnArgSerValIlePheSerSerGluArgSerGluGluG 406
579 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 539
406 ulysThrProAspAsnLeuThrSerGlnLeuGlnIleGlnLeuL 423
538 CTCTCTACTT.....CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 495
423 ysSerGlyArgLeuValLeuLysAspArgAlaValLeuSer..... 436
494 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 445
437 .....AlaProSerLeuSe 441
444 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 395
441 rGlnAspProGlnAlaLeuLeuLeuMetGluAlaGlyThrSerLeuLysT 458
394 TCTCCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 345
458 hrSerSerAspLeuLysLeuAlaThrLeuSerIleProLeuHis..... 472
344 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 295
473 .....SerLeuAspThrGluLysSerValThrIleHisAl 484
294 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC...TC 248

```

484 aProAsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspGluA 501

247 TCCCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 198

501 snPhetATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 517

197 CTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 166

518 Leu...LeuThrLeuSerLysGluGlnSerHisLeuHisLeuProAspG 533

165 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 116

533 yAsnLeuSerSer 537

115 CTCTCTCTCTCTCTCT 103

seq_name: qb_est76:BE636683

seq_documentation_block:

LOCUS BE636683 1183 bp mRNA EST 25-AUG-2000

DEFINITION rockefeller.0.373 Mastigamoeba balamuthi lambda Zap II Library

ACCESSION BE636683

VERSION BE636683.1 GI:9919794

KEYWORDS EST.

SOURCE Mastigamoeba balamuthi.

ORGANISM Mastigamoeba balamuthi

REFERENCE 1 (bases 1 to 1183)

AUTHORS Lee,J.A., Moore,D.V., Gordon,P., Sensen,C.W., Gaasterland,T. and Muller,M.

TITLE cDNA clones (expressed sequence tags) from the free-living

amitochondriate amoeboflagellate, Mastigamoeba balamuthi

COMMENT Unpublished (2000)

JOURNAL Contact: Muller Miklos

LABORATORY Laboratory of Biochemical Parasitology

ADDRESS The Rockefeller University

1230 York Avenue, New York, NY 10021, USA

Email: mmuller@rockefeller.edu

Insert Length: 1183 Std Error: 0.00

POLYA-No.

FEATURES

source

1. 1183

/organism="Mastigamoeba balamuthi"

/db_xref="taxon:30984"

/clone_lib="Mastigamoeba balamuthi lambda Zap II Library"

/note="syn: Phreatamoeba balamuthi"

alignment_scores:

Quality: 99.00 Length: 361

Ratio: 0.576 Gaps: 13

Percent Similarity: 47.645 Percent Identity: 20.222

alignment_block:

US-09-677-752-4 x BE636683 ..

Align seg 1/1 to: BE636683 from: 1 to: 1183

118 ProSerIleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnGln 134

199 CCGGCGACCTCTCTCGGCAAGGAGGCAAGCTGAGATCCCTCAACGG 248

134 Y.....SerMetSerPheCysArgAsnHisAlaGluLys 146

249 CCACGCTCTACACATCGCGCTCTCGGCGCTCAGAACCCCGGAGATCC 298

146 erGlyGlyAlaIleSerAlaAspAlaPheSerLeuGlnHisAsnGlyLeu 162

Intermodal stem segments"
 BASE COUNT 172 a 179 c 95 g 219 t
 ORIGIN

alignment_scores:
 Quality: 100.00 Length: 200
 Ratio: 0.870 Gaps: 12
 Percent similarity: 57.500 Percent identity: 28.500

alignment block:
 US-09-677-752-4 x AW694070 ..

Align seg 1/1 to: AW694070 from: 1 to: 665

```

359 SerSerSerSerSerSerProleuProleuSerSerSerSerSerSerVal17h 375
|||||.....|
7 TCTTCTCTATCCCATCACAACACCTTCAGATCGATCGCTCGCATCG 56
375 rSerProThProAlaThAlaSerProleuVal11egLThSerAlaA 392
|||||.....|
57 CAACATTCACCAATTCACCTCGCATTCATCATC..... 93
392 snArgSerVal11epheser.....SerLuargleuSerGluGlu 406
|||||.....|
94 .....GTGATTTCTCTCTCTGCTTCCCAATATCATCAACGAAGAA 135
407 LysThrProAspAsnLeuThrSerGlnLeuGlnGlnProIleGluLeu 423
|||||.....|
136 CGAAGCAGCAGCAGCTTCTTCCCTTCATCAAGATCC.....GA 173
423 sSerGlyArgLeuValLeuLysAspArgAlaVal..... 434
|||||.....|
174 TTCAGATATATCATCTCAACGACATCATCTCTGTTACTTACTT 223
435 LeuSerAlaProSerLeuSerGlnAspProGlnAlaLeuLeuMet 450
|||||.....|
224 CCATCAAAACCCCGA.....CCCTATTTCTCTCTATCTTC 261
451 GluAlaClyThrSerLeuLysThrSerSerAspLeuLysLeuAlaThrLe 467
|||||.....|
262 TTCACGCTACTCACTTCAATGACATGACATGACATCAACACCGAAGT 311
467 uSerIleProLeuHisSerLeuAspThrGluLysSerValThrIleHisA 484
|||||.....|
312 TCACAAGAAATTC...TCCATCGTCGCTTCTTCATTCATTCACCAACG 358
484 LArProAsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspGlu 500
|||||.....|
359 CTAACTCTTCACTCTTCAAAACTCTCTTTTGT.....GAATC 399
501 AsnPhetYrGluAsnValGluLeuLeuSerLysGluAsnAsnIlePr 517
|||||.....|
400 GAATTCAGAAATCG.....CAGCTCAGTTCTCTCAATTCGGTGTAA 443
517 oLeuLeuThrLeuSerLysGluInserHisLeuHisLeu.....ProA 532
|||||.....|
444 CGCTCTT.....CCCATATTCGCTCTCTGCTGCTCTA 475
532 spGlyAsnLeu.....SerSerHisPheGlyTyrGlnGlyAspThr 546
|||||.....|
476 ATCATGGGTTAGAGATTCGACATATGAT...CAAGGGGATTTCTCT 522

```

seq_name: gb_gss25:A2671472

seq_documentation_block:
 LOCUS A2671472 883 bp DNA GSS 14-DEC-2000
 DEFINITION ENTJUT77TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
 genomic, DNA sequence.
 ACCESSION A2671472
 VERSION A2671472.1 GI:11808738
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica.

ORGANISM Entamoeba histolytica

Eukaryota; Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 883)

AUTHORS Loftus,B., Van Aken,S. and Fraser,C.

TITLE Determination of clone end sequences from Entamoeba histolytica

JOURNAL HMI:IMSS sheared DNA library

COMMENT Unpublished (2000)

Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

Email: b.loftus@igf.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared

DNA library

Seq primer: M13-Forward

Class: shotgun

High quality sequence start: 31

High quality sequence stop: 867.

Location/Qualifiers

FEATURES

source

1..883

/organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: PHOS1; Site: 1; Bst I; Constructed at the

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica

using a method described by Clark and Diamond (Clark,

C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a

method for isolate identification. Exp. Parasitol.

77:450.). The DNA was mechanically sheared to give a

light size distribution (~2 kb). The v + i method used for

the library construction is described in detail in Smith,

H.O. and Venter, J.C. (Making small insert libraries for

whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaundin and B.

Barrell, Oxford University Press, 1999)."

122 c 133 g 260 t

BASE COUNT

368 a

ORIGIN

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

122 c 133 g 260 t

COMMENT

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LICM502 row: h column: 18
 High quality sequence stop: 605.
 Location/Qualifiers

FEATURES

source

1. 774
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3829697"
 /clone_lib="NIH_MGC_58"
 /tissue_type="hypertrophoma"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site: 1: SfiI (ggccgctcggcc); Site: 2: SfiI (ggccatctggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-ATTCTAGAGCGCCGCGCCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.35 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 89 a 297 c 101 g 287 t
 ORIGIN

alignment_scores:
 Quality: 101.00 Length: 247
 Ratio: 0.808 Gaps: 13
 Percent Similarity: 50.607 Percent Identity: 28.340

alignment_block:
 US-09-677-752-4 x BF030678

Align seg 1/1 to: BF030678 from: 1 to: 774

```

331 11etyleuglulysaspalaialeuserseleuglualaargangl 347
      ::::::::::::::::::::
75  cttcttctctctctctctctctctctctctctctctctctct 113
      ::::::::::::::::::::
347 yasp1leuepheasprolevalgin.gluuserseleuglu 363
      ::::::::::::::::::::
114  ....cttcttctctctctctctctctctctctctctctctct 156
      ::::::::::::::::::::
364 serpro.....leuproserseleuglualaservajthrserpr 377
      ::::::::::::::::::::
157 tcccttccgctctctctctctctctctctctctctctctctct 206
      ::::::::::::::::::::
377 othrproalathralaserproleuvalieglnthseralasnarg 394
      ::::::::::::::::::::
207 cttcttctctctctctctctctctctctctctctctctctct 256
      ::::::::::::::::::::
394 erval1lepheserleugluargleusergluglu1ysthproasp 410
      ::::::::::::::::::::
257 cttcttctctctctctctctctctctctctctctctctctct 306
      ::::::::::::::::::::
411 asmeuthrsergluenglnproilegluueuserseleugluargle 427
      ::::::::::::::::::::
307 tctcttctctctctctctctctctctctctctctctctctct 347
      ::::::::::::::::::::
427 uval1leulysaspargalavalleuseralaproserseleugl 444
      ::::::::::::::::::::
348 cttcttctctctctctctctctctctctctctctctctctct 373
      ::::::::::::::::::::

```

seq_name: gp_est89:BF570499

```

444 roglalaleuileumetglualaglythrserleugluthrser 460
      ::::::::::::::::::::
374 cttcttctctctctctctctctctctctctctctctctctct 411
      ::::::::::::::::::::
461 aspleu1leu1alathrleuserleleproleu1h1serleu1asp1h1 477
      ::::::::::::::::::::
412 gctgttctcggcttgcagctgcgcctctctctctctctctctc 461
      ::::::::::::::::::::
477 ulysseval1thrle1h1alapro1nleuserle.....g 490
      ::::::::::::::::::::
462 tccactagagggcagctggcattcattgttactctgttggttacacca 511
      ::::::::::::::::::::
490 1nlysl1lephleuserasnservglaspgluasnphethyrgluasnval 506
      ::::::::::::::::::::
512 agnaaatatcggggctcactgctggcagccaccacactcagacna..atc 558
      ::::::::::::::::::::
507 gluleu1euser.....lysglu1nasnasn1leproleu1thrle 521
      ::::::::::::::::::::
559 tccactgttaagctcccccataaaaacagcccttcattctctctc 600
      ::::::::::::::::::::
521 u1serlysglu1n1ser1h1leu1h1leup1roasp..... 532
      ::::::::::::::::::::
601 .....ccattcagccattgttatgttgcattgccatggcctcagaatcag 643
      ::::::::::::::::::::
533 ....glyasnleuserse1h1pheg1ytrg1ngly..... 543
      ::::::::::::::::::::
644 agtgtggamaac...tcnctacgcttggaatctctgtgagccggacac 690
      ::::::::::::::::::::
544 .....aspr1thrpheser1r1p1s 550
      ::::::::::::::::::::
691 gggctcttcagggctcgtgaacccctttctctcttgagaga 729
      ::::::::::::::::::::

```

seq_documentation_block:

LOCUS BF570499 1466 bp mRNA EST 12-DEC-2000
 DEFINITION 602075417f1 NIH_MGC_62 Homo sapiens cDNA clone IMAGE:4242700 5',
 mRNA sequence.
 ACCESSION BF570499
 VERSION BF570499.1 GI:11644211
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS NIH-MGC help://mgc.ncl.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: ATCC/DC/DTP
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LICM1052 row: a column: 05
 High quality sequence start: 21
 High quality sequence stop: 112.
 Location/Qualifiers

FEATURES

source

1. 1466
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4242700"
 /clone_lib="NIH_MGC_62"
 /tissue_type="melanotic melanoma, high MDR"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: skin; Vector: pDNR-LIB (Clontech); Site: 1: SfiI (ggccgctcggcc); Site: 2: SfiI (ggccatctggcc); Double-stranded cDNA was prepared from cell line RNA. 5'

Mon May 7 09:06:43 2001

us-09-677-752-4.rst

Page 12

```

565 .....LysAsnTyrValProHisProGluArgInSerThrLeuVal 578
      ::::| | | | | | | | | | | | | | | | | | | | | |
1158 CCGGACACTGACCTTCGCCATCATCTACACATCTCTGCCCTAATT 1206

```

```
seq_documentation_block:
  docnum: 82538438
```

LOCUS	883 bp	DNA	GSS	03-NOV-2000
DEFINITION	Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.			

SOURCE	Entamoeba histolytica.
Entamoeba histolytica	

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 803)	Loftus, B., Van Aken, S. and Fraser, C.	Determination of clone end sequences from <i>Entamoeba histolytica</i>

```
alignment_scores:      100.00    tooth.h: 313
```

```
alignment_block:
```

US-09-677-752-4 x AZ528428/rev

Align seg 1/1 to reverse of: A2528428 from: 1 to: 883

```

373  SerValThrSerProThrProAlaThrLaserProLeuValIleGlnTh 389
      ::::::::::: ::::: ::::: ::::: :::::
697  ACCGTACAGACATCATCCCGTTAAATATTGATGATGCCAACAGCTC  648
      ::::::::::: ::::: ::::: ::::: :::::
389  tGsrAlaAnaArgSerValIlePheSerSerClnArgLeuSerClnGlnG  406

```

```

647 TAAATTAATGACGTCAGTAGTAAAGACACCAAGCAAGAAAGATTGAA... 602
406 IuLysThrProGspAsnLeuThrSerGlnLeuGlnProIleGluLeu 422
601 .....ATACCCAAACCGAACAAATT 581
423 LysSerGlyArgLeuValLeuLysAspArgAlaValLeuSerAlaProSe 439
580 AAA..... 578
439 rLeuSerGlnAspProGlnAlaLeuLeuIleMetGluAlaGlyThrSerL 456
577 .TTACTGTTAGTAGACAGACGCTATTGTTCTATGAGAAAGACCAAAACCT 529
456 euLysThrSerSerAspLysLeuAlaIleThrLeuSerIleProLeuHis 472
528 CACTGCACAAAGAAAGAAC.....TGATGACAAATGAACTTTAC 491
473 Ser.....LeuAspThrGly 478
490 CAAAGCAACATGATTTTCAACCAATGACAGAAATTAAGCTTGATGCCAATA 441
478 sSerValThrIleHisAlaPro.....AsnLeuSerIleGlnLysI 492
440 AGATGTAGACTTTACTGTCATGTCATCATCATCAATCAATCATATGTTCA 391
492 lPheLeuSerAsnSerGlyAspGluAsnPheTrpGluAsnAlaGluLeu 508
390 TCATTCTAGCAAAAGGAGACCAATGACAAATTAATATACATTTGACAGA 341
509 LeuSerLysGluGlnAsnAsnIleProLeuLeuThrLeuSerLysGluI 525
340 TTA.....TATATTCTAATACTGCTCAAGTTAAACAGCAAGCTTCAGAA 297
525 nSerHisLeuHisLeuProAspGlyAsnLeuSerSerHisPheGlyTyrG 542
296 GTTAAACTTATTACCAACAGATGGGGATTCACAAACCAATTTATATATG 247
542 lncLysAspTrpThrPheSerTrpLysAspSerAspGluGlnHisSerLeu 558
246 ATGGAAGAAAATAACTCATTTATTATTCATGCAATTGTAGACAAACAATT 197
559 IleAlaSerTrpThrProLysAsnTrpValProHisProGluArgGlnSe 575
196 .....CAATTATATCAAAAAAATATCATT..... 173
575 rThrLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMetGlnAla 552
173 ..... 173
592 aGlnSerMetIleAsnThrIleAlaHisGlyAlaTyrLeuPheGly 608
172 .....TCGTGTCACATGGTGATTTGCTACTTAACCTACC 140
609 ThrTrpGlySerAlaValSerAsnLeuPheTyrAlaHisAspSerSerG 625
139 ATGAAGAACGTCGATCACAATTAATATTATAATGACAGAACTTAATTG 90
625 yLysProIleAspAsnTrpIleHisAspSerLeuGlyTyrLeuPheGly 642
89 TACTGATATG.....TACA 76
642 lSerThrHisSerLeuAspAspHisSerPheCysLeuAlaAlaGlyGln 658
75 TTACCACTGAAGAAGTATAAA.....TGTTTAAATGTGGGAGAA 38
659 LeuLeuGlyLysSerSerAspSerPheIleThrSer 671
37 .....GGAAAAACCGATTAGGAGACATGCTTTCCAGCACA 5
req_name: qb_gss23:R2548511

```


[illegible]

seq_documentation_block: 1048 bp DNA GSS 17-MAY-2000

LOCUS CNS03JGH Tetraodon nigroviridis genome survey sequence 77 end of clone

DEFINITION 03JC18 of library G from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION AL246842

VERSION AL246842.1 GI:7967854

KEYWORDS GSS: genome survey sequence.

SOURCE Tetraodon nigroviridis.

ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinoptera; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Eurypterygii; Ctenosquamata; Acanthomorpha; Euacanthomorpha; Holacanthopterygii; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 1048)

AUTHORS Roest-Crolius,H., Jaillon,O., Dasilva,C., Fzames,C., Fisher,C., Bouneau,L., Billault,A., Querier,F., Saurin,W., Bernot,A. and Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

JOURNAL unpublished

REFERENCE 2 (bases 1 to 1048)

AUTHORS Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,

[illegible]

```

/db_xref="taxon:4513"
/clone="HV_CED0017K08f"
/clone_lib="Hordeum vulgare seedling green leaf EST
library HVGDNA0004 (Erysiphe infected & control)"
/tissue_type="seedling green leaf"
/lab_host="TJCL121"
/Note="Vector: lambda2AP; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      261 a      743 c      35 g      599 t      30 others
ORIGIN

alignment_scores:
    Quality: 105.00      Length: 406
    Ratio: 0.590      Gaps: 19
    Percent Similarity: 43.842      Percent Identity: 22.660

alignment_block:
US-09-677-752-4 x BF064737

Align seg 1/1 to: BF064737 from: 1 to: 1688

349  ILeuPheAspProIleValGlnGluSerSerLysGluSerPr 365
      |||||.....:|||||
596  ATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 639
      :|||:|||||
365  oLeuProSerSerLeuGlnAlaSerValThrSerProThrProAlaThrA 382
      |||||.....:|||||
640  CTTTCATCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 680
      :|||:|||||
382  IAspProLeuValIleGlnThrSerLysAlaSerValIle.PheSe 398
      :|||:|||||
681  CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 718
      :|||:|||||
398  rSerGluArgLeuSerGluGluGluLysThr..... 408
      |||||.....:|||||
719  ATCATCTCATCTACTACCATTCATCTACTATCATCTCATTCATNCATCT 768
      :|||:|||||
409  .....ProAspAsnLeuThrSerGln 415
      :|||:|||||
769  CATCTCTCTCTCTCTCATATCATTCCTCTCTCGCCATCTCTCTCCCACT 818
      :|||:|||||
416  LeuGlnGlnProIleGluLeuLysSerGlyArgLeu..... 427
      |||.....:|||||
819  CTATCTCTCTCTCTCTCATTCACACTCTCTCTATCCCTACTCTCTCTC 868
      :|||:|||||
428  .....ValLeuLysAspArgAlaVal 434
      :|||:|||||
869  ACATCATATCTCTCATATCTACTTCTATTCATCTCTCTCTCTCTCTCA 918
      :|||:|||||
434  AlLeuSerAlaProSerLeuSerGln.AsprProGlnAlaLeuLeuIleMe 450
      :|||:|||||
919  CTCTACACTACTCATCTCATCTCATCTCTCTCTCTCTCTCTCTCTCTCTCT 968
      :|||:|||||
450  tGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAlaThrL 467
      :|||:|||||
969  ATCATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1018
      :|||:|||||
467  euserLleProLeuLysSerLeuAsp..... 475
      |||.....:|||||
1019  ATTCTCTATCTCTATTCATCTCTACTCTCTCTCTCTCTCTCTCTCTCTCC 1068
      :|||:|||||
476  .....ThrGluLysSerValThrIleHisAl 484
      :|||:|||||
1069  TCTCTCTCACTACTACCATCTCTCTCACTCTCTCTCTCTCTCTCTCTCTCA 1118
      :|||:|||||
484  aspAsn.....LeuSerIleGlnLysIlePheLeuSerA 496
      :|||:|||||
1119  TCACACTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1168
      :|||:|||||
496  snSerGlyAspGluAsnPheTyrGluAsnValGluLeuLeuSerLysGlu 512
      :|||:|||||
1169  CATCCACACCTTTCTCATCTCTATTCGTACTCTCTCTCTCTCTCTCTCTCT 1218

```

```

513  GlnAsnAlaIleProLeuLeuThrLeuSerLysGluGlnSerHisLeuH 529
      :|||:|||||
1219  TCTTCACATCTATTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1253
      :|||:|||||
529  sleuProAspGlyAsnLeuSerSerHisPheGlyTyrGlnGlyAspTrrp 546
      :|||:|||||
1254  CTAC.....TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1284
      :|||:|||||
546  hrPheSerTrpLysAspSerAspGluGlyHis.SerLeuIleAlaSerTr 562
      :|||:|||||
1285  .....CACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1300
      :|||:|||||
562  pThrProLysAsnTyrValProHisProGluArgGlnSerThrLeuValA 579
      :|||:|||||
1301  ACACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1349
      :|||:|||||
579  IAsnThrLeuTrpAsnThrTyrSerAspMetGlnAlaValGlnSerMet 595
      :|||:|||||
1350  CATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1372
      :|||:|||||
596  ILeasnThrIleAlaHisGlyGlyAlaTyrLeuPheGlyThrTrpGlySe 612
      :|||:|||||
1373  TCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1407
      :|||:|||||
612  rAlaValSerAsnLeuPheTyr.....A 620
      :|||:|||||
1408  TAGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1457
      :|||:|||||
620  IAsnAspSerSerGlyLysProIleAspAsnTrpHisAspSerLeu 636
      :|||:|||||
1458  GTCATCTCCATATCATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1507
      :|||:|||||
637  GlyTyrLeuPheGlyIle.....SerThrHisSerLeuAspAspH 650
      :|||:|||||
1508  CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1557
      :|||:|||||
650  sSerPheCysLeuAlaAlaGlyGlnLeuGlyLysSerSerAspSerP 667
      :|||:|||||
1558  TTCACATATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1582
      :|||:|||||
667  helleThrSerThrGluThrSerTyrIleAlaThrValGlnAlaGln 683
      :|||:|||||
1583  .....ACTGCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1627
      :|||:|||||
684  LeuAlaThrSer 687
      :|||:|||||
1628  CTCACACACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1639
      :|||:|||||

```

seq_name: gb_est85:BF279812

```

seq_documentation_block:
LOCUS      BF279812      1660 bp      mRNA      EST      17-NOV-2000
DEFINITION GA_Eb0033P14 Gossypium arboreum 7-10 dpa fiber library Gossypium
ARBOREUM CDNA clone GA_Eb0033P14, mRNA sequence.
ACCESSION  BF279812
VERSION    BF279812.1 GI:11210882
KEYWORDS   EST.
SOURCE     Gossypium arboreum.
ORGANISM   Gossypium arboreum.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Malvales; Malvaceae; Gossypium.
REFERENCE  1 (bases 1 to 1660)
AUTHORS   Wing,R.A., Fritsch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
D., Wood,T.C., Leslie,A. and Wilkins,T.A.
TITLE     An integrated analysis of the genetics, development, and evolution
of the cotton fiber
JOURNAL   Unpublished (2000)
COMMENT   Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288

```

531 oaspglyasnleuserSerHisPheGlyTyr 541
 174 C.....CTTCTCTCCATCATCATCTCTAT 150
 seq_name: qb_gss17:A2218282

seq_documentation_block:

LOCUS A2218282 576 bp DNA GSS 09-JUN-2000
 DEFINITION Sheared DNA-81G10.TF Sheared DNA Trypanosoma brucei genomic clone
 Sheared DNA-81G10, DNA sequence.

ACCESSION A2218282
 VERSION A2218282.1 GI:8436082

KEYWORDS GSS.
 SOURCE Trypanosoma brucei.
 ORGANISM Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 576)

AUTHORS El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Sub, E., Malek, J., Fujii, C., Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J., Fraser, C. and Adams, M.

TITLE Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library

JOURNAL Unpublished (1999)

COMMENT Other_GSSs: Sheared DNA-81G10.TF

Contact: Najib M. El-Sayed
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

Email: nelsayed@tigr.org

Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through Research Genetics, Alabama, USA. Sheared DNA end sequences search page: <http://www.tigr.org/cdb/mdb/cddb/>.

Seq primer: M13-Reverse

Class: shotgun.

FEATURES

SOURCE

1..576 Location/Qualifiers

/organism="Trypanosoma brucei"
 /strain="TREU927/4 GUTat 10.1"
 /db_xref="taxon:5691"
 /clone="Sheared DNA-81G10"
 /clone_11b="Sheared DNA"
 /note="Vector: pUC18; Site.1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + 1 method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 119 a 132 c 121 g 204 t
 ORIGIN

alignment_scores:

Quality: 106.50 Length: 224
 Ratio: 0.772 Gaps: 11
 Percent similarity: 61.607 Percent identity: 21.875

alignment_block:

US-09-677-752-4 x A2218282 ..

Align seg 1/1 to: A2218282 from: 1 to: 576

285 IlaasnAsnSerAlaLysIleGlyAlaIleAlaIleGlnSerGlyG1 301
|||||.....
 6 TTGCAATCATCATCTTTCAGTCACTGATGTTGCAATCATCGGT 55

301 ySerLeuSerIleLeuAlaGlyGlnGlySerValLeuPheGlnAsnAsn 318
 |||||
 56 TTCATGCAGTCACCTGTGATGTCATCATCTACG 90
 318 eGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeuGlu 334
 :||
 91 TTTCATTTGAGTCACCTGTG...ATGTTCAATCATCATCGTTTCATTTGACG 137
 335 LysAspAlaIleLeuSerSerLeuGlnAlaArgAsnGlyAspIleLeu 351
 :|||
 138 TCACCTGTGATGTCATCATCATCATCTTTCATTCACCTGTGATGTCGCAATCA 185
 351 ePheAspProIleValGlnGlnSerSerSerGlnSerProLeuPro 368
 :|||
 186TTCATCATCATCGTTTCATTCACCTGTG... 218
 368 eTyrLeuGlnAlaSerValThrSerProThrProAlaThrAla... 382
 :|||
 219 ..ATGTTGCAATCATCATCATCTTTCATTCACCTGTGATGTCGCAATCA 266
 383SerProLeuValIleGlnThrSerAlaAsn...Ar 393
 :|||
 267 TCGGTTTCATTCAGTCACCTGTGATGTCATTCATTCATTCATTCATTCGCA 316
 393 gSerValIlePheSerSerGlnArgLeuSerGlnGlnGlyThrProA 410
 :|||
 317 GTACCTGTGATGTCATCATCATCATCGTTTCA..... 347
 410 sPAsnLeuThrSerGlnLeuGlnGlnProIleGlnLeuLysSerGlyArg 426
 :|||
 348TTGCACTCACCTGTGATGTCATTCATTCATTCATTCATTCATTCGCA... 377
 427 LeuValLeuLysAspArgAlaValLeuSerAlaProSerLeuSerGlnAs 443
 :|||
 378GTTTCATTCAGTCACCTGTGATGTCATTCATTCATTCATTCATTCGCA... 409
 443 pPro.....GlnAlaLeuLeuIleMetGlnAlaGlyThrSerLeuL 457
 :|||
 410 ATCGGTTTCATTCAGTCACCTGTGATGTCATTCATTCATTCATTCATTCG 459
 457 yThrSerSerAspLeuLysLeuAlaThrLeuSerIleProLeuHisSer 473
 :|||
 460 ACTCACCTGTGATGTCATTCATTCATTCATTCATTCATTCATTCATTCGCA 500
 474 ..LeuAspThrGlnLysSerValThrIleHisAlaProAsnLeuSerI 489
 :|||
 501 CCTGTGATGTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCGCA 550
 489 eGlnLysIlePheLeuSerAsn 496
 :|||
 551 ATCATCGTTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCGCA 572
 seq_name: qb_est93:BF864174

seq_documentation_block:

LOCUS BF864174 1166 bp mRNA EST 19-JAN-2001
 DEFINITION 963046G03.Y1 C. reinhardtii CC-1690, stress condition I, normalized
 , lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BF864174
 VERSION BF864174.1 GI:12254318

KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii.

ORGANISM Chlamydomonas reinhardtii.

Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 1166)

AUTHORS Grossman, A., Davies, J., Federspiel, N., Harris, E., Hauser, C., Lefebvre, P., McDermott, J.P., Shlager, J., Sillflow, C. and Stern, D.

TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in

Vascular Plants: project phase 3

JOURNAL Unpublished (2000)

```

410 ASPAspLeuThrsGlnLeuGlnGlnProIleGluLeuLysSerGlyAr 426
    :|||:|||||
628 GTTCTTGTCTCTCT.....TCTGTAC 650
426 GLeuValLeuLysAspArgAlaValLeuSerAlaProSerLeuSerGlnA 443
    :|||:|||||
651 ATTGTG.....GATGTAATGTCAGTCTCTTCTGCCACATGTT 668
443 sPrGlnAlaLeuLeuIleMetGluAlaGlyThrSerLeuLysThrSer 459
    :|||:|||||
689 CACCA.....TTCTGCCCATCTTCCACACACAGCATCTCTCTCTCC 729
460 SerAspLeuLysLeuAlaThrLeuSerIleProLeuHisSerLeuAspTh 476
    :|||:|||||
730 TCTTCTCAGCTCTTCTCTCT.....TCTTCCAAACATCTTCGGCTGTAC 776
476 r...GluLysSerValThrIleHisAlaPro.....AsnLeuSerIleG 490
    :|||:|||||
777 ATCAGAAAGATCCGCTCTGTCTCTCTCCACGATCATCTCTGTCTTC 826
490 InLysIlePheLeuSerAsnSerGlyAspGluAsnPro 502
    :|||:|||||
827 CTCTCTAGATTGCTCTCTCTGATGATGCTTCCTCTTC 864
seq_name: gb_gss23:A2545202

```

```

seq_documentation_block:
LOCUS      A2545202      874 bp      DNA      GSS      14-NOV-2000
DEFINITION ENTDR55TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
ACCESSION  A2545202
VERSION    A2545202.1  GI:11165757
KEYWORDS   GSS.
SOURCE     Entamoeba histolytica.
ORGANISM   Entamoeba histolytica.
REFERENCE  1 (bases 1 to 874)
AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
TITLE     Determination of clone end sequences from Entamoeba histolytica
          HM1:IMSS sheared DNA library
          Unpublished (2000)
JOURNAL    Contact: Brendan J Loftus
COMMENT    Department of Eukaryotic Genomics
          The Institute for Genomic Research
          9712 Medical Center Dr., Rockville, MD 20850, USA
          Tel: 301 838 0200
          Fax: 301 838 0208
          Email: b.loftus@tigr.org
          Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
          DNA library
          Seq primer: M13-Forward
          Class: Shotgun
          High quality sequence start: 17
          High quality sequence stop: 831.
          Location/Qualifiers
            1..874
              /organism="Entamoeba histolytica"
              /strain="HM1:IMSS"
              /db_xref="taxon:5759"
              /clone_lib="Entamoeba histolytica Sheared DNA"
              /note="Vector: pHOSt1; Bst I; Constructed at The
              Institute for Genomic Research (TIGR), Rockville, MD.
              Genomic DNA isolated from broth cultures of E. histolytica
              using a method described by Clark and Diamond (Clark,
              C.S., and Diamond, L.S. (1993) Entamoeba histolytica: a
              method for isolate identification. Exp. Parasitol.
              77:450.). The DNA was mechanically sheared to give a
              tight size distribution (~2 kb). The v + i method used for
              the library construction is described in detail in Smith,
              H.O. and Venter, J.C. (Making small insert libraries for
              whole genome shotgun sequencing projects. In Genome
              Sequencing: A Practical Approach, eds. M. Vaubin and B.
              Barrell, Oxford University Press, 1999)."

```

```

BASE COUNT      332 a      99 c      136 g      307 t
ORIGIN
alignment_scores:
  Quality: 112.00      Length: 392
  Ratio: 0.636      Gaps: 17
  Percent Similarity: 44.898      Percent Identity: 22.194
alignment_block:
US-09-677-752-4 x A2545202/rev
Align seg 1/1 to reverse of: A2545202 from: 1 to: 874
155 PheSerLeuGlnHisAsnThrLeuPheThrAlaPheGluGluAsnSerSe 171
    |||:|||||
854 TTATATATAAACCTTATATTTA.....GAACATTATTC 820
171 rLysGlyAsnGlyGlyAlaIleGlnAlaGlnThrPheSerLeuSerArg 188
    |||:|||||
819 TATCTTAACCAAAACCAATTTTATTTCTTCTCCATTAATTTTAACA 770
188 snValSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnGlyGly 204
    |||:|||||
769 ATCTTCTCTATTTACATTTGCTCATCTTACATCT.....CTTAT 732
205 AlaIleCysCysSerAsnLeuIleCysSerGlyAsnValAsnProLeuPh 221
    |||:|||||
731 .....TGCTGTGAATTAGTCTCTCTCT.....CTTAT 703
221 ePheThrGlyAsnSerAlaThrAsnGlyAlaIleCysCysIleSerA 238
    |||:|||||
702 TTATGATTCAAAT.....GTAAGAACTTTAGTG 674
238 sP.....LeuAsnThrSerGluLysGly 245
    |||:|||||
673 ATTATTCATTAGTGGATTGGAACATTTGTAAGATTGCTCTTAAGGC 624
246 SerLeuSerLeuAla.....CysAsnGlnGlnThrLeuPheAlaSerAs 260
    |||:|||||
623 TCAATTTAATACAGCTGTGCAATCT.....600
260 nSerAlaLysGlyLysGlyGlyAlaIleThrAlaLysHisMetValLeuA 277
    |||:|||||
599 .....GAATGATGATGAATTAATATTCACACAAATTTATATATTT 560
277 rGlyAsnGlyProValSerPheIleAsnAsnSerAlaLysIleGlyGly 293
    |||:|||||
559 CATAT...TGCTCTGTAGTGTGTCACGATGTTCTATTAAATACCAAC 513
294 AlaIleAlaIleGlnSerClyLysSerLeuSerIleLeuAlaGlyGlyG 310
    |||:|||||
512 ACTATTTCAAGAGTACTGCTCTAGTATCAAGTATTCAGTATTCAGGA 468
310 ySerValLeuPheGlnAsnAsnSerGlnArgThrSerAspGlnGlyLeu 327
    |||:|||||
467 .....AAATTAATTA 458
327 aArgAsnAlaIleThrLeuGluLysAspAlaIleLeuSerSerLeuGlu 343
    |||:|||||
457 TCAATGGGATCTTCATTAAGTAAATTTAACCCTAATTTCTACATTAACA 408
344 AlaArgAsnGlyAspIleLeuPheAspProIleValGlnGluSerSe 360
    |||:|||||
407 ATTACAGTCAATCTACTGTAATTAATGAAGCACACAGTCACTCAAT 358
360 rSerLysGluSerProLeuProSerSerLeuGlnAlaSerValThrSerP 377
    |||:|||||
357 AAATAGTATTAATCATCATCTGATCTATTT.....C 326
377 rThrProAlaThrAlaSerProLeuValIleGlnThrSerAlaAsnArg 393
    |||:|||||
325 CAATTTAATCTTCAATCACTACATTAATTAATGAGCGTTAAATAA 276

```

212 TCACAGTGGAGAGCTTCTTGACACACTGCTTGTCCCTGACACAGCAT 261
 383 erProleuValIleGlnThrSerAlaAsnArgSerValIlePheSer 399
 262 CTCCTATTATGTCACCAACCATGACCAACAGTATTATGACCATGCA 311
 400 GlnArgLeuSerGlnGluGlyThrProAspAsnLeuThrSerGln 416
 312 CACCAACAACAACACACAATAGCTCTT.....CAATATAT 349
 416 uGlnGlnProIleGlnLeuLysSerGlyArgLeuValLeuLysAspArg 433
 350 TCACACACCA..... 359
 433 laValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeuIle 449
 360CGACGACCTTCT.....CAACCAACAATATATACAGT 392
 450 MetGlnAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAla 466
 393 CCAAGGTGATGATCATCTGCACGACATCATCT.....GCAAC 430
 466 rLeuSerIleProLeuHisSerLeuAspThrGlnLysSerValThrIle 483
 431 ATCTACTCAACCA.....C 444
 483 laValProAsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAsp 499
 445 AATGCCACAATCATATGCGCAGTATTACTATCATTAATACCAATGAC 494
 500GluAsnPhenylGluAsnValGlu.....LeuLeu 510
 495 AATAAAGAAAGAAACTTCTGAGAGAAAGAAACAATGCTGTGAGAA 544
 510 rLysGlnGlnAsnAsnIle 516
 545 AAGAGACAAACGATATT 563
 seq_name: gb_gss32:CNS05015
 seq_documentation_block:
 LOCUS CNS05015 1086 bp DNA GSS 26-MAY-2000
 DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
 03H114 of library B from Tetraodon nigroviridis, genomic survey
 sequence.
 ACCESSION AL348697.1 GI:8242467
 VERSION AL348697.1
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Eurypterygii; Ctenosquamata; Acanthomorpha; Eucanthomorpha;
 Holacanthopterygii; Acanthopterygii; Perciformes;
 Tetraodontiformes; Tetraodontidae; Tetraodon.
 1 (bases 1 to 1086)
 Rost-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
 Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
 Weissenbach,J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 Unpublished
 2 (bases 1 to 1086)
 Rost-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,
 Saurin,W. and Weissenbach,J.
 Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 Unpublished
 3 (bases 1 to 1086)
 Genome.
 TITLE JOURNAL
 REFERENCE AUTHORS
 TITLE Direct Submission

JOURNAL Submitted (12-Apr-2000) to the EMBL/GenBank/DBJ databases
 COMMENT This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.
 FEATURES
 source
 1.1086
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone_id="034114"
 /clone_lib="B"
 /note="Genoscope sequence ID : C0NB034BE07C1-end : T7"
 BASE COUNT 311 a 241 c 239 g 282 t 13 others
 ORIGIN
 alignment_scores:
 Quality: 116.00 Length: 300
 Ratio: 0.758 Gaps: 17
 Percent Similarity: 51.000 Percent Identity: 25.000
 alignment_block:
 US-09-677-752-4 x CNS05015 ..
 Align seg 1/1 to: CNS05015 from: 1 to: 1086
 384 ProleuValIleGlnThrSerAlaAsnArgSerValIlePheSerGln 400
 105 CCATATTATTAAATCAAGTCGAATGACAAACGATGTTGACAGCCC 154
 400 uArgLeuSerGlnGlu.....GlnLys.ThrProAsp 410
 155 TTTTGTCTATTTCAGACTCTCGCGCGCGCTGTGAAGAACCATGGCC 204
 411 AsnLeuThrSerGlnLeuGlnPro...IleGlnLeuLys...SerG 425
 205 CGGTTGGATGCGCTGCCAAGAAAGATTATTGACCTCGGGAATCTGG 254
 425 yArgLeuValLeuLysAspArgAlaValLeuSerAlaProSerLeuSerG 442
 255 TTGAGATTTCCTAAATCAACGACATCTCTCGAGCTCGAGACATTAAAG 304
 442 lAspProGlnAlaLeuLeuIle..... 449
 305 TGTCTCGCAGCGCAATCATCTCTCTCGAGAGATTTCATGAGACCA 354
 450MetGlnAlaGlyThrSerLeuLysThrSer 460
 355 CACGAGAGAGTACGCCCTGTGAGGCTGGAAGC.....ACCTCGTC 395
 460 rAspLeuLysLeuAlaThrLeuSerIleProLeuHisSerLeuAspThrG 477
 396 ATCAGACGACGATGCAM.....CCTCAGTCGGAACATCAGCAG 433
 477 lLysSerValThrIleHisAlaProAsnLeuSerIleGlnLysIlePhe 493
 434 GAAGAGAGAGCAATATTTCACGTTCCAAACATTGCCAGCACTTCAT 483
 494 LeuSerAsn...SerGlyAspGluAsnPhenylGluAsnValGluLeu 509
 484 CATGTAACCTTCTTGAACAGGAAATTTACAGACACGCTTACATT 533
 509 uSerLysGlnGlnAsnIleProLeuLysThrLeuSerLysGlnGln 526
 534 TTCGAGAGCCGATGCAATTTTCC.....ACAAAGGAAGCA 571
 526 erHisLeu...HisLeuProAspGlyAsnLeuSerSerIlePheGly 541
 572 GCGAGCTGGAACCAACATTAATGAATAAAGAGAAAA...GGTAT 617
 542 GlnGlyAspTrp..... 545
 618 TCAGATTGTTGGCGTCACCTCTTTCGACAGAACACAAATAATATC 667

Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of *Chlamydia trachomatis* MOpn and *Chlamydia pneumoniae* AR39.
A:Reference number: AB1500; MUID:20150255
A:Accession: BB1593
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-978 <REF>
A:Cross-references: GB:AE002191; GB:AE002161; NID:q7189216; PIDN:AA38155.1; PID:q7189227
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0298

```

Query Match 583; 11.5%; Score 583; DB 2; Length 978;
Best Local Similarity 26.1%; Pred. No. 5,6e-28;
Matches 263; Conservative 137; Mismatches 411; Indels 196; Gaps
45

Oy 41 SNKISLTGDTNHL--TNCYL-DNLRYIALLOKTPNEGAAVTTIDLSFDFTOKECIYR 97
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 67 SNNLSLKGKGLSLFTSCQAPNTSNVLL-----SAEELT-FKFN-----SSINF 112
Oy 98 KNLTFEESGALICYASPNSEYVEIROTIGVIFENN-----TCORPFSNPNAAVNRIG 153
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 113 GMSOTGLGLLI-YGMD-----IVFOSIKDLITFTTNVAVSPASVTTSAIP--AITYVTG 164
Oy 154 G-----AHHQNTLYINHHNDVGFEMKFEYSYVGGAISTANTFVVS--ENOSCELFMDN 204
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 165 ASALOPTDSLTYEN--ISOSIKFEGULANF-----GSAISSPVAAYKFINNTATMFSFN 218
Oy 205 ICIOTNTAGKGAIVAGTNSNFSNNCDLEFINNACC-----AGCAIFS 248
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 219 F-----TSSGGGYIYGGSSILFEENSGCIIETANSVNSLKGVTPSSGYALAGSGLCI 273
Oy 249 P--ICSLTGCRCNIVIVNNRCFKNVETASSEASDGAIKWTRIDLVYGNRGRIFFSDNIT 306
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 274 PIGTFELKNQCKTTSYN-----CTPDACAIVAEI--CNVNGGALLDSNTA 322
Oy 307 KNYGAIYAPVYTLVDNGPTYFINNIANNKGAIIY-----DGTSNKISADRIAIT 358
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 323 ARNGGAIKAKVLNIOGRPIEFESRNRA-EKGGAIFIOPSYDPAKOTSTLLTILASEGNIA 381
Oy 359 FNNIYNTNTNAGTGTANAPRRNRIYVSSSGEILIGASSQNLIIFYDPI-----E 411
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 382 FQGNML-----NTRPGIRNALTVEAGGEIYSLSQGSRVLYFPIIHSLEPTS 450
Oy 412 VSNAGYVSFNKEADOTGYSVYFSGATVNSAD-FHORNLOTKTPAPLTLNSGFLCIEDHAQ 470
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 431 PSKKDITINN---GASGSVYFTSKSLSTELLPAVTTIILLGTIVKIASGELKITDNV 487
Oy 471 LTVNRF--TOTGGVYSLGNGAVLSCKYKGAQNSASMAITLKHIGLNLSSILKSAEIPLL 529
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 488 VNVLGATQOSGCLTLGSGTLGLATPTGAPAA--VDFTIGKLAFDPSFLKRD----- 539
Oy 530 WVEPTNNSNYTADLTATFSLDVKLSLIDYGNSPPESTDLTALSSQPLMSISEASDN 589
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 540 FVNASVWAG--TKNVTLTGAL-----VDE-----HDVIDLVMVSLQSPVAIPAV-- 584
Oy 590 QLRSDMDQFSGLVN-----PHYWGQCLMTGWAK--TQDEPASSAATITDPQ 635
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 585 -----FKGATVTKTGFDPDEGIATPSHHGQYOKWSTYSRPLIPADPGFGPGSPS 636
Oy 636 ANRHFRTLLTW-----LPAGVYSPHRSPLRLANTLWGNMLLTATESLKNA-----EL 684
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 637 AN-----TLIYAVNSDTLVRSTYIIDPERGEGIVNSLM-----ISFLGNAPFSDIIO 686
Oy 685 TPSDHPWGTIGGGLGMNYYODPRENHPGFHMRSGSYSAGIACOTH--TFLSKFSQTYT 742
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 687 LLIDHPGLSTAKALGAVVENTPRQGEHGFSGRAGGYOALSMNVTDTHTTLGLSGOLYG 746
Oy 743 KLNERYAKNNVSSKNY--SCQGENLFLSDEQFLLTKLVGLSYGDHNCNHFYTOGENT--T 799
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 747 KTNANPYDSRCESEMYLLTFEGPPIYTOKEALISWKAAYGYSKNHNLNTTYTLRPDKRPK 806

```

QY 800 SGGT:RSOTMGAVFEDLPKMFSTHILTP-----FLALGYSSLSHFTEVGA 850
Db 807 SGGQHHNSVYVLISAE---HPLWKCCLTLPRLAQMDSLFIAEFLGQMSQFTETGD 863
QY 851 YPFSFTKPLTNLVLPVIGVKGSENNATOR-POAMVELAOPVLYROEPCITQALLASK 909
Db 864 LQRST-RGKGYNSLPLGCGSQWTFPKKAPSTLTITLAKLPRIYRNPINITYVSNQ 922
QY 910 GIMFSGSGSPSSRHAMSYKLSOOTQPLSMVLTLHFQYHGFYSSSTFCNY 956
Db 923 ESTSISGANLRHNGLFVQI-HDVVDLTEDTQALFLUYTFDQKNGFNNH 968

RESULT 24
F72076
Polymorphic outer membrane protein g family - Chlamydophilia pneumoniae (strain CML0239)
C.Species: Chlamydophilia pneumoniae, Chlamydia pneumoniae
C.Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C.Accession: F72076
R.Katman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999
A.Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A.Reference number: A72000; MUID:992060606
A.Accession: F72076
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-973 <ARD>
A.Cross-references: GB:AF001629; GB:AF001363; NID:g4376734; PIDN:AAD18595.1; PID:g4376734
A.Experimental source: strain CML029
C.Gene(s):
A:Gene: pmp_13

Query Match	11.5%	Score 575	DB 2	Length 973
Best Local Similarity	25.5%	Pred. No. 1.7e-27		
Matches 273	Conservative 139	Mismatches 390	Indels 268	Gaps 50
Qy	1	MKKAFFFLIGNSLGLAREVPSRITLMPNSVDPPTKESLSNKI	---	SLTGDTHLT-- 54
Db	1			
	1	MKTSIRKFLISTT	---	LAPCFASTATVEVIMSENFDDSSCGKIFPYTLTSDPRGLCIF 57
Qy	55	---NCYLDNRLRYLIALIIOKT	---	PRGGAAYITTDVLSFFDQKEGIEYPAKNL-PRGSGAIG 109
Db	58	SGGLYIANDLMDNALSRTSSSCFNSRAGALQLG	-----	KGVGFELNRSADCAAIS 109
Qy	110	YASPNSTPVEIRDTTG	---	PVIFENNTCCRPFTSSNPNAAVNKIRREGAIIHAQULYNHNH 167
Db	110	SVTTONPELCPLSFSGFSQMIFDN	---	CESLTJD--TSASNVLPHASAIATTPMLEFTNN 164
Qy	168	DVYGFKNFS	---	YVREGAISTANTFVNSENOSCFLENDN----- 204
Db	165	DSLIFYQNRGACGAMIRGTSITIENT	-----	KSLLFPGNCSISNGCALTGSAAILNI 218
Qy	205	-----ICIQNTNG	---	KGALY-AGTSNFSNNODLFETINACCGAIFSPICSLTGNR 257
Db	219	NNSAPVLEFSTNAGIYGALYLTGSGMLTSGNLGVLNVNNSRSRGAITA	-----	N 270
Qy	258	CINVEFYNNR	---	CEKNYETASSEAS-----DGGAIKYV--T 288
Db	271	GNVTSNNSDLITGN	---	NTASPONSLPAPTPPTPPATVPLGGVGAIFCIPATPTPTGV 329
Qy	289	RLDVTGNRGRIFFSDNITKNYGAIYAPVYTLVDNGPTT	---	FINNIANNKGAITYIDGTSNS 348
Db	330	SLTISGNS	---	VTFELINASECGGLAKKISISNKEITFLGNA--CKGAIALAPEGEL 387
Qy	349	KISDRHAIFENETIVNTNANGTSTAMPNRRNATVASSCEILLGAGSSONLIFYD		408
Db	388	SLSANODILEFNKL	-----	SLTSGPTTNSHFGKDAFALGALOGITLIFYD 437
Qy	409	P1	---	EVSNAGYVSFKNREAD-----QTGSVYFSGATVNSADFHORNLOTKTPA----- 454
Db	438	PTSDDLTASASAAATVYVNPKASADGAYSGTIVSGEFLTAT	-----	EAATPANATST 490

A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: D81601
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1,947 <REA>
A:Cross-references: GB:AE002182; GB:AE002161; NID:q7189140; PIDN:AE38083.1; PID:q718914
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: pmp_19; CP0213

Query Match 11.5%; Score 586; DB 2; Length 947;
Best Local Similarity 25.2%; Pred. No. 3,5e-28;
Matches 250; Conservative 173; Mismatches 390; Indels 178; Gaps 47;

14 LSGLAWEPSRIELMPNSVDPPTKESLSNKI---SLTGDTNLTMCYLDNRYLAILQK 70
Db LSGIHSGEDLEFLTRSSSPTKTYSLRKDFIVCDPAGNSIIRPGAAFLNLGDFEINS 89
Qy 71 TPNEGAAYITDYLSPFDQKGIYFAKULTPESGALIGYASPNSTPVE-----IRPT 123
Db 90 TPL--AALTFKNI--HLGARGAGLSESNTV-----FKGLHSLVLENNESWGCVLTT 137
Qy 124 IGPVIFENNT--CCRPETSSNPNAVANKIREGAIHAQN-----LYINHHVDVGFMKNF 176
Db 138 SCDLFLINNTSVLQCNINISYGP-----GGALLQGRKSKALFEFRDNRGFI LFLKMK 188
Qy 177 SYVR-----GGAISTANTFVYSENOGCFLEMDN--TCIQNTNACKGAIYAGT--SNSE 226
Db 189 AVNODESHGPGYGAVSSISP-----GSPITPADNOELFQENEGELGAIYNDGALITE 242
Qy 227 ESNNDLFLINNACGAGAFSPICSLTGNRGNIYVNNRCKNVTASSEASDGAIRV 286
Db 243 ENNFOTTFSSKASFGGAVISRYCULYSQKDTLF-----TKNAAKVGCAIR- 291
Qy 287 TRLDVTGNRGRIFESDNTTKNGAIYAVYTLVD--NGPTVFINNIA--NNKGAIYIDG 344
Db 292 ADVVHTRDCKGSLVFEEN--SATAAGAIAVNAVCIDINAGPVRFINNSALGLNGAIYMQA 350
Qy 345 T--SNSKISADRAITFENENIVTNV--TNANGTSTANPRRNAIIVYASSGCEILLGAGSS 401
Db 351 TGSILRLHANOQDIEFCGKNKVRQSFHSHINSNFTN---NAITIGAPREFSLSANEG 406
Qy 402 ONLIFDPI-----EVSNAGVSFENKADDTGVSFGATVNSADRHORNLQTKTP---A 454
Db 407 HRICTPDITISATENTNS--LYINHQRLEAGCAVIFSGARLSE--HKENKKTISIINO 463
Qy 455 PLTSLGFLCIEDHQAOLTVNRFTOTGGVNSLNGAVLSCKYKNGNSASNASITLKHIGL 514
Db 464 PVALGCGVLSIEGAILAARFEGGGLLALGSGSKLTT---QGNKSEKDKIVITNLGF 519
Qy 515 NUSILKSG--AEPLLMVEPTNNSNNTADTATFSLSDVKL-----SLID--DVGNSE 565
Db 520 NLENDLSSDPAET-----RATEKASIEISGVPVRYGHTSEFEYENHEKASKR 565
Qy 566 YESTDTHALSSQPMPLSIEASDNOLRS---DDMDFSGLVNPHYGOGGLTWMQATQDP 622
Db 566 YTTISIT---LSAKKLVTAAPRKQDIONLLIASEYWG-----YVQGSMEFNSNDTK 617
Qy 623 EPPSSAATIDPOKARNFHRTLLTLWLPAG--YVSPKHSRSLPLANTLGNML-----LA 674
Db 618 E-----KTIIASMTPTGEFSLDPKRGSGFIPTLWSTFGSLINISIV 661
Qy 675 TESLKSASLETBDHPWMTGGGLGMVYQDRENNHPCGHMSSSGSACMIAGQHTFS 724
Db 662 NNNYLNNSSEVILPQH--LCVFGGPVYOIMEONKROSSNNLLVOHAGHNG--ARIPEFN 717
Qy 735 LKRSQYTKLNEYAKNNVSKNSYSCGEMLF---SLQEGFLTKLVGLVSYGDHN---C 788
Db 718 TIIISALITOLFSSSSQOONVADKSHA---QILIGTVSLNKWQALSLRSSSTYEDSDYMK 774
Qy 789 HNFYTOGENTLSQGTFRSQTMGAVFFDLPMKDFGTHILTAFLALGIYSSL--SHT 846

Db 775 HVFPYK---TSRGSWRNYGWSGYSGMAY--PKGIRFLKMTFPVLDQ--YTKLVQNEFV 828
Qy 847 EVGAYPRSESTKTPILNVLVPGVKGSEFMNATOPQAM--TVELVQPLVROEGEATQL 905
Db 829 ETGIDPRFSS--SEMTNLSLPIGIALEMRIFGSKSLFLOYSTYIKDLRVRVQSSASL 887
Qy 906 LASKGIMFGSGSPSSSRHMS-----YKI 928
Db 888 VLNHTWDIQCPVPLGKEALNITLUNGSTIKYKI 918

RESULT 21

B72077

polymorphic membrane protein G family CP0306 (imported) - Chlamydia pneumoniae (s
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000

C:Accession: B72077; B81592
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999

A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606

A:Accession: B72077
A:Molecule type: DNA

A:Residues: 1-928 <RNA>
A:Cross-references: GB:AE001628; GB:AE001363; NID:q4376730; PIDN:AD18591.1; PID:q437

A:Experimental source: strain CWL029
R:Reed, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255

A:Accession: B81592
A:Molecule type: DNA

A:Residues: 1-928 <REA>
A:Cross-references: GB:AE002192; GB:AE002161; NID:q7189226; PIDN:AE38163.1; PID:q718

A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: pmp_9; CP0306

Query Match 11.5%; Score 585.5; DB 2; Length 928;
Best Local Similarity 24.9%; Pred. No. 3.7e-26;
Matches 266; Conservative 139; Mismatches 380; Indels 285; Gaps 46;

1 MKKAFEPFLIIGNSL-----SGLAREV-----PSRIPLMNSVDPD--TKESLSNKIS 45
Db 1 MKSSLHMFILSSSLPLSLNSAFPAAYVEINLGPINSFSGGCTTPPAQTNNAGCTIYN 60

Qy 46 LTGDTNLTN-----CYLD--NURYI---LAILQKTPNEGAAYTID----- 82
Db 61 LTGADV--SITNAGSPALTPASCREFETGNLSFOHGVOFLQNDAGANCTINFAANKLL 119

Qy 83 -----YLSFDTQKRGITFAKULTPESGALIGYASPNSTPVEIRDTQPIYFENNTCR 136
Db 120 SFEGSYSLDIQT-----TNATGTGA-----IKSTGACSIQSNISCF 158

Qy 137 PFTSNPNAVANKIREGAIHAQNLVYINHHVDVGEKNSYVRGCAISTANTFVSENO 196
Db 159 GONFSNDN-----GALOGSSISLSLNPNT--FAKKAKAQKAGALSTGGITINNTL 209

Qy 197 SCFLFMNDNICIQNTAGKAGAIYAGTSNFSNCDLFIINNACCA-----GGAISPTCS 252
Db 210 NSASFSEN-----TAANNQGAITYEAS--SFISSNKAISFINNSVTATSAITGAIY---CS 260

Qy 253 LTGNRGNIYFNNRCKNVTASSEASDGAIVTTRLDVTVGNRGRIFESDNTTKNYGA 312
Db 261 STS-----ADRPVLTLS-----DNGELNLTGNTAITSGGA 290

Qy 313 IYAPVTVLNDGPTVFINNIANK-----GGAIVIDGTSNKSISADHAAIFENIVTNT 368
Db 291 IYDNLVLSGGFTLLKKNNSAIDTAPPLGAIYIADSGSLSLALOGDITFEGNTVYKKA 350

A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0302

Query Match 11.7%; Score 597; DB 2; Length 949;
Best Local Similarity 26.1%; Pred. No. 7,4e-29;
Matches 253; Conservative 148; Mismatches 361; Indels 206; Gaps 47;

```
OY 36 TKESLSKISLITGD-----THNLNLCY-----LDNLRI-----LAILQKTPMEGA 76
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 66 TPKISATTVSLTGQVFFPEPKGPISDSCKPOTDMLTLCNGSLTFCPIADACTHAGA 125
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 77 AVTTTDLSEFDQKEGICFPAKNTLPESGCAIGVASENSPVEIRDTIGPIFENNCCR 136
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 126 AASTT-----ANKNLTFSGFSLSDSSPSTTV-----TTGQ----- 157
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 137 PTTSSNPAANVKIREGGAIAQNLINHHHDVYGFKNFSYVGAIASTANTFVSENO 196
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 158 -----GTLSSAGVLENI-----RKLIVAG-----NFTADGCAIKGA-SPLLTGTS 199
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 197 SCFLPHNICIQTNTAKGCAIYAGTSNFSNNCDL--FFINNACCAGAIIFSPICSLT 254
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 200 GDAIFSNK-----SSSTKGGAI-ATTAGARIANTGYVRELSNIASTGSAIDDEGTSTL 253
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 255 GNGCNIFYNNRCFKNVETASSEASDCAIKVTTRLD-----VTGNGRIFESDNTKNG 310
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 254 SNKKFLFEEN-----AAKTTGCAICNTKASGSPELLISNNKTLIFASNVAETSG 303
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 311 GAIVAPVTVLDNGPTYFI--NNI--ANNKGAIIYIDGTSNKSISADRAHAIIFENIVTNV 367
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 304 CAIHAKKLALSSGGEFTFLRNWVSATPKGCAISIDASGELISAETGINITFVNTLT-- 361
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 368 TMANGISTANPPRRNAIYVASSSGELLGAGSSONLIFDPI--EVSNAQVSSEFKEA 425
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 362 -----TTGSTDTPTKRNAININGSKFTELRAKANHITFFPDPTSEGTSDVILKINNCSA 416
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 426 -----DOTGSVVFSGATVNSADFH--QRNLQRTYPAFLTLNSGFLCIEHQAOLTVRFTQ--T 479
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 417 GALNPYGTILFSGEITLADBLKADNLKSFTQPVSLSGKLLQGVLESTFSFGQEA 476
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 480 GGVVSLCNGAVLSCYKNGAGNSASNASITLKHICLNLSILKSGAEPLLWVEPTNSNN 539
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 477 GSLIGMOSGTTL-----STTAGSITITLINGINVDSL--GLKQPV-----S 514
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 540 YTDNTATFSLSDVKLSLIDYGSPEESDTLHALSSQPLMSISEASDNLRSDDMDFS 599
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 515 LTAAGASNKVIVSGKLNLIIDEGNI--YES-----HMFSDQLFSLKITVDADVDTNDIS 569
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 600 GL--NVP-----HYGMOGLMTWMAKTODEPASSATITPOKANRRHRTLLTLWLPAGY 652
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 570 SLIPVPAEDPSEVEGFGOGMNVNN-----TTDTATNTKEAT-----AATWITGCF 613
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 653 VPSSKARHSPLIANTLMGNMLLATESLKSASAE--TPSDHP--FWGITGGCLGMWVDDPR 708
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 614 VPSPERSALVCLNTLMG--VFTDIRSLQOLVEIGATGMEHKGFV--VSSWNTNLFHKTGD 669
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 709 ENHGFIMHRSRGYAGMLAGOTHT-----PSLKSQTYTKLNERY--AKNN----- 752
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 670 ENRKGFRHTSGY--VIGGSAHTPKDDLFFAFCHILPARKKDCFLAHNNSRTYGGTLFF 726
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 753 -----VSSKNYSOCGEMFLSLQ--EGF-----LFTKLVLGYISYGDHCHHFTOGENLTS 800
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 727 KHSHTLPQONVLRBGRAKFSESAIEKPREIPLADVOVSSHNDMMETHTYSLPE--S 784
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 801 OGTRSQTMGCAVFFDLPMKPFSGTHIITAFPLGALG--IYSSLSHFTEGAYPRSEST 857
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 785 EGSMSNECIAGICLIDLPF-VLSNPHPLFKFIFIPQMKEMVYVSONSEFESSSGRGFSI 843
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 858 KTLPIINVLPVIGVSGFEMNATOROPAMTVELA--YQPLVQEOEGIAITOLLASKGIWFGS 915
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 844 GR-LLNLSIPVGA--FVQG--DIGDSTIYDLSGFVSVYVANNPOSTATILVMSFDSKIR 899
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
```

OY 916 GSPSSRHA 923
DB 900 GGNLSRGA 907

RESULT 18

H71460
Probable outer membrane protein H - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
R:Accession: H71460
R:Stephenson, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A:title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A:Reference number: A71570; MIMD:99000809
A:Accession: H71460
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1016 <ARN>
A:Cross-references: GB:AE001360; GB:AE001273; MID:g3329342; PIDN:AA68470.1; PID:g332

A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: pmpH

Query Match 11.7%; Score 595; DB 2; Length 1016;
Best Local Similarity 25.4%; Pred. No. 1.1e-28;
Matches 261; Conservative 149; Mismatches 404; Indels 214; Gaps 49;

```
OY 42 NKISLTDGTHNLT-----NCYLNLRYLIALIQKTPMEGAAVTITDYI.SFFDQKEGIYF 96
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 69 DNLITTGQNHLSPTDSQGVQLQVAFISA-----GETLTLKDRSS-----LWF 112
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 97 AKNLTPESGCAIGVASENSPVEIRDTIGPIFENN-----CCPR 137
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 113 SKNVSCGEKGM-----SGKTVSISGA-GEVIFWDSNGVSPLSIYVASTPRPAPAPAP 166
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 138 FTSSNPAANVKIREGGAIAQNLINHHHDVYGFKNFSYVGAIASTANTFVSENO 197
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 167 AASSLSPTVSDARKSIFSEVET-----SLEISGVKKGVNF-----DNMG 207
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 198 CF--LEFDNICIQTNTAGKCAIYAGTSNFSNNCDL--LFTINACAGAIIFSPICSL 252
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 208 NFGTVFRCGN--SNNAGSGGS--GSATTPSFYKNCXKGVSFIDNVAACGCGVYKGTVL 263
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 253 LTGRCGNIFYNNRCFKNVETASSEASD-----CGAI-KYTRLDVYGNGRIFESD 303
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 264 FKDEGGIFFRGNTAYDDIGLIAATSRDONTETGGCGGVCSPDDSVKFPBNGKSTVFDY 323
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 304 NITKNYGAIVAPVTVLDNGPTYFIINNINANKGAIY--IGTNS-----KISAD 353
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 324 NFAKGRGSLITKEFSLVADSVFNSNNTAKKGAIIATITDSTGGSILERRRAAE 383
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 354 RHAIIFNE-----NIVNTVNA-----NCTSTSANPPRRNAITVASSSGEILLGAGSSO 402
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 384 GGAICVSEASGSGTGNLTLSASDGIYFSGNMTSDRGERSAARILSDGTIVSLNAGLS 443
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 403 NLIFYDPIEVSN--AGVS-----VSFNKEADQVGSVVSAGATVNSADFHQ--RN 447
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 444 KLIFYPDVQVNSAAGASTSPSSSSMPGAVTTINQSN--GSVIFTAESLTPSEKIQVLN 501
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 448 LOTKTPAPILNSNGFLCIEHQAOLTVRFTQGTGVVSIAGCAVISCYKNGAGNSASNASI 507
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 502 STSNFPAALIVSGGELVTVGATITLTITATSGRVTLGSGASLSAAGANN--NYTC 558
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 508 TLKHTIGLNSILKSGAEIPLLWVEPTNSNNYTAATATFSLSDVKLSLIDYGSNPE 567
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 559 TVSKLIGLDESFLTPNPKTAIILGADGT-----TVVNSGSTLDT--VWESAEVYDN--PLF 610
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 568 STDTHALLSQPMLSTSEAS--DNOLRSDMDFGSLVPHPHGMOGLMTWMAKTODEPAS 626
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 611 VGSILT-----IPVTVLSSSSASNGSVTKNSVTTINDADAHAHYGOGSMAQMTK--PPLAP 662
    | : : : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
```

Db 948 VRGSMYAS-----LTGNIIEYGC 964

RESULT 14

A:1591
polymorphic membrane protein G family CP0307 (imported) - Chlamydia pneumoniae (strain
C:Species: Chlamydia pneumoniae pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: A81591
R:Read, T.D.: Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: A81591
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-930 <REV>
A:Cross-references: GB:AE002193; GB:AE002161; NID:g7189234; PIDN:AF38164.1; PID:g718923
C:Genetics:
A:Gene: CP0307

Query Match 12.2%; Score 620; DB 2; Length 930;
Best Local Similarity 27.1%; Pred. No. 2.8e-30;
Matches 252; Conservative 139; Mismatches 374; Indels 164; Gaps 40;

47 TGDTHNLNCLYLDNRLYLALLOKTPNEGAATV---ITDYLSEFDTOKEGIFYAKNLTP 103
Db 50 TADA-NGTNVYLSGNVYI-----NDACKGTALTGCCFETTGDLTFGKGYSFSFN-TVD 102
Qy 104 SGGAIGVSPSPYIEIRDT-----IGPIYFENNTCCRFSTSSNPAANVKREGCAIHAQ 159
Db 103 AGSNAGAASATADKALFTGFSNLSFIAAPGTVAASGKSTLSAGALNTLDNCTILFESQ 162
160 NLYINHNDVYGFKNFVSVRGCAISTANFVSENOGCFLEPMONICLOTNTAGK-GGAI 218
Db 163 NVSNENANN-----GGAI-TAKTILISGNTS-----SITFTSNKAKLGGAI 203
Qy 219 YAGTNSFESNNDLFFINN-ACCAGAI-FSPICSLGTNRGNIVFYNNRCKNVTASS 276
Db 204 YSSAASISGNTGQLVPMNNKGTGGALGFEEASSITON-SLFEFGN-----TATD 255
Qy 277 EASDGAIV-----TTRLDVTGNRGRIFESDNTITKNNGCAIYAPVTVLVNCGPTTYINN 331
Db 256 AAGKGAIVCEKTEGTPTLTISGNKS-LTFAENSSVTGGALCAHGLDLSAAGPTLFSNN 314
Qy 332 IANN-----KGAIVYIDGTSNKSISADRAHAIFFENIYNTNANGTSTANPPRRNATV 387
Db 315 RCGNTAAKGAIALADSGSLSLANOGDITFLGNTL-----TTSAPSTRNATVYL 366
Qy 388 ASSSGEILLGAGSSONLIFYDPIEVSNAGVS-----VSFNKEADOTGVSFVGATVNS 440
Db 367 GSSAKITMLRAAOCOSIFYDPIASNTTGASDVLTINOPDSNPLDYSGTIVFSGEKL-S 425
Qy 441 ADFHRR-NLOCTPAPPLTSLNGFLICEDHAOLYNNRTOTGGVSLONGAVLSCTYKGA 498
Db 426 ADEAAADNFTSILKOPPLASGLTALKGNVELDNGFTOGESTLL-----MQGT 477
Qy 499 GNSASNAITTKHIGLNSILKSGAEIPLVWEPTNNSNNVTADTAATFSLSDVKLSLI 558
Db 478 KUKADTEAISLKLKLVVDLSAL-----EGNKSVSITAGAKNTITLTS-PLVFO 524
Qy 559 DDCNSPYESITDLTHALSSQPLSLISEASDNLRSDDMDFSGLVN-----PHYGWL 611
Db 525 DSSGNF-YESHTINQAF-TOPLVFTAAAT---AASDIYIDLALTSPTQTEPRHYGYGH 578
Qy 612 WTGCAKATQDEPPASSATITDPKARNRHRLLTLWLPAGVYPSKHKHSPLIANTLKNM 671
Db 579 WEATWA---DTSTAKSGT-----MTWVTTGYNPDPERRASVAVPSLMAASF 620

Qy 672 LLATESLKNASAEIPESDHPFW---GITGGGLGMVYODPRENHGPFHMRSSGYSAGMIAG 728
Db 621 ---IDIRTLQOIMTSQANSIYOQKGLANSGTANFEHKKDSGTNOAFRHSIGYIVGSSAE 677

Qy 729 --QHTFSLKFSQTYTKLINERYAKNNVSSKNVSCO-----GEMFSLQ 769

Db 678 DFSENIFFSVARCOLFGKDKDLFIYEN-TSHNVYLSLYLQHRAPFLGGLPMPSPGSIIDMLK 736

Qy 770 EGFLLTKLVGLYSYCDHCHHFFYT-----QGENLTSQSTFSSQTMGCAVFFDLPMK-PFG 823

Db 737 DIPLINQOLSYSTKNDMDTRYTSYPEAQSMTNNSGALD--LGSLALYLKPEAF 793

Qy 824 STHILTAFLGALGIVSSLSHFTVEGAVPRFSRTKPLINVLVPIGVGFSFNNATOPQA 883

Db 794 QGYF--PFLFQAVYSNQONFKESGACARAFD-DGDLVKNSTIVGIRLEKISEDEK-NN 848

Qy 884 WTVELAYQPVLYROEPGIATQLLASKGIW 912
Db 849 FEISLAVIGDYVYRKNRPSRTSLVMSGASM 877

RESULT 15

D72078
polymorphic outer membrane protein g family - Chlamydia pneumoniae (strain CWL029

C:Species: Chlamydia pneumoniae pneumoniae, Chlamydia pneumoniae

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000

C:Accession: D72078
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lamell, C.; Fan, J.; Ollinger, L.; Grimwood,

Nature Genet. 21, 385-389, 1999

A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A:Reference number: A72000; MUID:99206606

A:Accession: D72078

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-930 <ARN>

A:Cross-references: GB:AE001627; GB:AE001363; NID:94376721; PIDN:AD18590.1; PID:9437

A:Experimental source: strain CWL029

C:Genetics:
A:Gene: pmp_8

Query Match 12.1%; Score 617; DB 2; Length 930;
Best Local Similarity 27.0%; Pred. No. 4.3e-30;
Matches 251; Conservative 140; Mismatches 374; Indels 164; Gaps 40;

47 TGDTHNLNCLYLDNRLYLALLOKTPNEGAATV---ITDYLSEFDTOKEGIFYAKNLTP 103
Db 50 TADA-NGTNVYLSGNVYI-----NDACKGTALTGCCFETTGDLTFGKGYSFSFN-TVD 102

Qy 104 SGGAIGVSPSPYIEIRDT-----IGPIYFENNTCCRFSTSSNPAANVKREGCAIHAQ 159
Db 103 AGSNAGAASATADKALFTGFSNLSFIAAPGTVAASGKSTLSAGALNTLDNCTILFESQ 162

Qy 160 NLYINHNDVYGFKNFVSVRGCAISTANFVSENOGCFLEPMONICLOTNTAGK-GGAI 218
Db 163 NVSNENANN-----KGAIVYIDGTSNKSISADRAHAIFFENIYNTNANGTSTANPPRRNATV 387

Qy 219 YAGTNSFESNNDLFFINN-ACCAGAI-FSPICSLGTNRGNIVFYNNRCKNVTASS 276
Db 204 YSSAASISGNTGQLVPMNNKGTGGALGFEEASSITON-SLFEFGN-----TATD 255

Qy 277 EASDGAIV-----TTRLDVTGNRGRIFESDNTITKNNGCAIYAPVTVLVNCGPTTYINN 331
Db 256 AAGKGAIVCEKTEGTPTLTISGNKS-LTFAENSSVTGGALCAHGLDLSAAGPTLFSNN 314

Qy 332 IANN-----KGAIVYIDGTSNKSISADRAHAIFFENIYNTNANGTSTANPPRRNATV 387
Db 315 RCGNTAAKGAIALADSGSLSLANOGDITFLGNTL-----TTSAPSTRNATVYL 366

Qy 388 ASSSGEILLGAGSSONLIFYDPIEVSNAGVS-----VSFNKEADOTGVSFVGATVNS 440
Db 367 GSSAKITMLRAAOCOSIFYDPIASNTTGASDVLTINOPDSNPLDYSGTIVFSGEKL-S 425

Db 771 --ANSYFSSMGLAFTEVEFGR-SKDYVVCRSNHHACIOSVYLSTKOA-----LCGSYLF 822
Oy 784 GDHNCHEHYTOG-ENLTSOGTFRRSQT-----MGAVFEDLPKMPRGSTHIL-APF 832
Db 823 GDATIRASVGFQNGHMKTSYTFEESDVMDNMCVLGEIGVLPITVTPSKLYLNLBRPF 882
Oy 833 LGALGITYSLSHFTVEGATPRSFSTKTPILNVLPIGVGSGFMNATOPROAMTVELAYOP 892
Db 883 VOAEFSYADHESEFTEGDARAFRS-GHLMNLSPVGVAFEDRCSS-HPNKYSFMGATYC 940
Oy 893 VYRDEPGIATOLLASKGIMFGSGSPSRHMASYKISQOTOPLSWLTILHFOYHG 946
Db 941 DAVRTISGIGTQLLSHQEWTTTDAFHLARHGVIVGSMVAS-----LTSNIEVYG 990

RESULT 11

C72078
polymorphic outer membrane protein g family - Chlamydomonada pneumoniae (strain CWL029)
C.Species: Chlamydomonada pneumoniae, Chlamydia pneumoniae
C.Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C.Accession: C72078
R.Katman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A.Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A.Reference number: A72000; MUID:99206606
A.Accession: C72078
A.Status: Preliminary
A.Molecule type: DNA
A.Residues: 1-936 <ARN>
A.Cross-references: GB:AE001627; GB:AE001363; NID:g4376721; PIDN:AA018589.1; PID:g437672
A.Experimental source: strain CWL029
C.Genetics:
A.Gene: pmp_7

Query Match 12.7% Score 645; DB 2; Length 936;

Best Local Similarity 27.0% Pred. No. 8.2e-32; Mismatches 146; Indels 172; Gaps 44;

Matches 262; Conservative 146; Mismatches 392; Indels 172; Gaps 44;
Oy 75 GAAVTTDYLSFDTOKEGIFPAKNTLPESGAIGY-----ASPNPTVEIR 121
Db 54 GTTYSLSDVSPFNAGALGIPLASGCFLEAGDGLTFQGNHAKFAFINAGSAGTVAAT 113
Oy 122 DTGPIVFEKN-----TCRPFSSNPAAVANKIREGAIHAOUYLNHNHNDVGFPM 174
Db 114 SAADKNLLEFDFSRSLISCPSSLSPGOCALKSV-----GNLSLTGNSQII-FTQ 164
Oy 175 NFSYVRCGATSTANTFVSENOSCFLEMDNICIOTNTAGCAIYA-GTNSPESNNCDL 233
Db 165 NFSSDNGVINTKN-FLLSGTSGFASFSRN---QATFGKGGVYVATGTTTINSFGIYS 220
Oy 234 FFINNACAGCAIFS-PICSLTGNRGNIVFYNNRCRKNVETASSEASDGAIKVTT--R 289
Db 221 FSONLAKGSGALYSTDNCSITDN-FQVIFDGNSEMAAO-----AOGGAICTTDTKT 273
Oy 290 LDVTGNRGTFEEDNITKNKGAIVAVTVLVNGPTFYFINNANK-----GGAIVIDG 344
Db 274 VTLTGKN-NLSTFNNTALTGALISGLKYSISAGPTLFGFNSISGSGAGGGGALMIAS 332
Oy 345 TSNKISADHAIIFENITVNTNANGTSTSANPPRRNATVYASSSGEILLGAGSSQNL 404
Db 333 AGELALSATSGDITFNNOVTN-----GSTST-----RMLNIIIDTKAVSTRATQSI 382
Oy 405 IFYDPLEVSNAGSVS-----FNKEADOTGVYFSGATVNSD-FHQRRLQTKTP 433
Db 383 YFVDPI--TNPGRASTDTLNLNLADANSEIEYGAIVFSGEKLSPREKAIANVSTIR 440
Oy 454 APULTSNGFLICEDHAGLTNRFQTCGVVSLGCAVLSCYKKNAGASNASASTLTHIG 513
Db 441 QPAVLAKGDLVLRGVTVTFKDLTQSPGSRILMDG-----GTTISAKKANISLNGLA 492
Oy 514 LNLSSILKSGAEIPLLVWEPTNNSNNYTAATFSLSDVKYLSLIDYGNPSPESTDLTH 573
Matches 262; Conservative 146; Mismatches 392; Indels 172; Gaps 44;

RESULT 12

Db 493 VNLSL-----DGTNKALKTEADKINISLST-IALIDTEG-SFEYNNHKS 538
Oy 574 ALSSQPMKISISASN-QLRSDDMFSGLNRP--HYNGOGLTWGMATODREPASSATI 630
Db 539 A-STYPLLELTITAGANGTTTIGALSTLLOEPETHYGGMMQJSMAN-----ATSKI 591
Oy 631 TDPOKANHFTLTWLPACGYVSPKHSRPLIANTLGMNMLATESKNSAEILPSDHP 690
Db 592 GS-----INMRTGIIPSPERKSNPLNSLNGN-FIDIRSNOLLETSSSEP 638
Oy 691 F-----WGITGGGLGMVYODPRENHPGHRSSGYSAQINACQHTSEFLKFSQTYTKNE 746
Db 639 FERELW-----LSGIANFYRDSMPTNRCFRHISGVALGITTATPAEDOLTRA--FCQLFA 693
Oy 747 RYKKNVSSKN-----YSCGEMLSLDEGFL-----LTKLVGL----- 780
Db 694 R-DRNHITTKNKGIDTYGASLFFHNTGFLDI-ANFLMKATRAPVLSISQIPLSDA 751
Oy 781 ---YSGDHNCHNFYTOGENTLSQCTFRSQTGCAVFFDLPMKPRGSTHIL--TAPFLGA 835
Db 752 KRSYLTIDNHMKTYT--DNSIIKGSWRMDACDGLASLRF-VISVPLYLKEVEPFVKV 808
Oy 836 LCIYSSLSHFTVEGATPRSFSTKTPILNVLPIGVGSGFMNATOPROAMTVELAYOPVL- 894
Db 809 OYIYAHQODFERHAEGRFAFN-KSELIINVEIPGVY---FERDSKSEKTYDLTLMYILD 864
Oy 895 -YRDEPGIATOLLASKGIMFGSGSPSRHMASYKISQOTOPLSWLTILHFOYGFSSSTF 953
Db 865 AVRNRKCCSTSLASDANMAVGTNLAROGFVRAANHFQVNPIMETGOF-AFEVRSSS 923
Oy 954 CNVLGEIALRF 965
Db 924 RNV-NTNLSKRF 934

RESULT 12

polymorphic outer membrane protein g family CP0308 (imported) - Chlamydomonada pneumoniae (s
C.Species: Chlamydomonada pneumoniae, Chlamydia pneumoniae
C.Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C.Accession: B81591
R.Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hicke
/ C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A.Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A.Reference number: AB1500; MUID:20150255
A.Accession: B81591
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-936 <REA>
A.Cross-references: GB:AE002193; GB:AE002161; NID:g7189234; PIDN:AA018165.1; PID:g718
A.Experimental source: strain AR39, HL cells
C.Genetics:
A.Gene: CP0308

Query Match 12.7% Score 644; DB 2; Length 936;

Best Local Similarity 27.0% Pred. No. 9.4e-32; Mismatches 146; Indels 172; Gaps 44;

Matches 262; Conservative 146; Mismatches 392; Indels 172; Gaps 44;
Oy 75 GAAVTTDYLSFDTOKEGIFPAKNTLPESGAIGY-----ASPNPTVEIR 121
Db 54 GTTYSLSDVSPFNAGALGIPLASGCFLEAGDGLTFQGNHAKFAFINAGSAGTVAAT 113
Oy 122 DTGPIVFEKN-----TCRPFSSNPAAVANKIREGAIHAOUYLNHNHNDVGFPM 174
Db 114 SAADKNLLEFDFSRSLISCPSSLSPGOCALKSV-----GNLSLTGNSQII-FTQ 164
Oy 175 NFSYVRCGATSTANTFVSENOSCFLEMDNICIOTNTAGCAIYA-GTNSPESNNCDL 233
Db 165 NFSSDNGVINTKN-FLLSGTSGFASFSRN---QATFGKGGVYVATGTTTINSFGIYS 220
Oy 234 FFINNACAGCAIFS-PICSLTGNRGNIVFYNNRCRKNVETASSEASDGAIKVTT--R 289

Db 652 WQSPHNLFLATLRYOTQOQIAPT-----ASGEATRLFLVHONSNDANGFHEATGSLG 705
Oy 725 MIA--GGHTFSLKFSOTYTKLNERAKNNYSKNSCGEGELFS-LOEGFLTLKVLGY 781
Db 706 TTSNTASHNSHSGVNSQLFNSLIEHSDNSVASHHTTVALQINNWLDERFSTSLA-X 764
Oy 782 SYCDHCHHFFYTOGEN--LTSOGTFRSOTMGAVFEDLPMKPFSGTHILTAFLCALGTY 839
Db 765 YSYN---HHIKASGVSGKIQTEGKCYSTLGLALSCSLQ- WSRPLHFFPFIQAIAYR 820
Oy 840 SLSLHTEVGAIPRFSFTKPLINLVPIGVKSGFMNATQORQATVETLQAPVLYXQEP 899
Db 821 SNOTAFQESGDKARFYSYHKKPLYNLTVPLGIQSAWESKFRPLTYWNIELAYQPVLYQONP 880
Oy 900 GIATOLLASKGIWFGSGSPSSRHAMSKYISQOTQPLSMTLTFLFOYHGFYSSTFCNYLNG 959
Db 881 ETVNSLESSGWSLSTGLTARNATAFKGRNOIFLPLKSLVLDYOGSVSSSTTHYLHA 940
Oy 960 EIALRF 965
Db 941 GTTFKF 946
RESULT 8
F71460
probable outer membrane protein F - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: F71460
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis. Reference number: A71570; MUID:99000809
A:Accession: F71460
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1034 <ARN>
A:Cross-references: GB:AE001360; GB:AE001273; NID:93329342; PIDN:AAC68468.1; PID:9332934
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: pmpF

Query Match 17.0%; Score 866.5; DB 2; Length 1034;
Best Local Similarity 28.1%; Pred. No. 2.3e-45;
Matches 293; Conservative 149; Mismatches 406; Indels 193; Gaps 36;

Oy 55 NCYLONLRYILAILLOKTPREGAAVTTIDYLSFFDTQKEGIFYFAKNLTPESGAI----- 108
Db 57 NIVLSMLQ-----SNGTGACTISG-----NTQTQFNSNVTNTADSGAFDMVTTTS 102
Oy 109 -----GYASPSPTVEIRDITIGPVFEENNTCCRPPTSSNPAAV-----NK 149
Db 103 PTASDANALLFCNNYCTHNGGGAIRSG-GPIRFLNNODVLPYNNISACAKYVGTGDHNE 161
Oy 150 IREGAIIHONLYINHHNDVYGFPMKFSYVYRGGAISTANTFVVSSENOGSLFMDNICI-- 207
Db 162 KNNGKALVYATITLIGNR-TLAFINMMSGCGAISADQISTIDIVKGLFENNHTLHN 220
Oy 208 ----QTNAGKGAITYAGTS-NSFESNCDLFFINNACGAGAIFFPISGLTNRGNIYF 262
Db 221 IPTYQENNAKGAICSRDLCSISNNSGPIYENYNOGKGCAISATRCVIDNNKERIIF 280
Oy 263 YNRRCKKNETASSEASDGAIVTTRLDVYGRGRIFESDNTKNYGAIVAPVTVLV 322
Db 281 SNNSSLG--WSOSSASNGCAIQTTOGFTLRNKKGSTIFPDSNATNAGCALNCIYDIRD 338
Oy 323 NGPTTFINNANKGAIYIDGTSNKSISADRAH-----IIFENIVTNTNANG 372
Db 339 NGPVYFLNNSA--AMGAAF-----NLKPRSATINYIHTGTDIYVNNNVFTLDGNL 389
Oy 373 TTSANPRRNALTIVASSSEIILGAGSSONLIYDPIE-----VSNAGV-----V 419

Db 390 GKKRLPHINNNEITPYTIS-----LGAKKDTRIYFYDLEQWERVKNENTSNNPSPTSRNTI 445
Oy 420 SENKEADQSVESGATVNSADFHQRNLQTK-----TPADLTLSNGFLCIEDHAQTLVN 474
Db 446 TVNPEFESGAVVVS-----YQMSSDIRFLMKENHYIKAEPTTLKFLGALIEDDAELEIF 502
Oy 475 R-FTQT-GGVVSLGNCNVLISCYKNGACNSASNSITLKHGLSLSLKSGAELPILMY 531
Db 503 NIPFTQNTSLLAGSGATLVGKHG-----KLNTINLGVLLPILKKGKSPCIRV 554
Oy 532 EPTN-NSNNYATD-AAEFSL-----DVKLSLDDGNSPEYSTDTHALSSQPLSI 583
Db 555 NQDDTQNTGIGTGPSSSTISSTIPMIIFNGRSLYDENVESYSDMSLRKACQLISTI 614
Oy 584 SEASDQLRSD--DMDFSGLVNPHYGMQGLMTWMAKT-----ODPEPASSATITD 632
Db 615 ETTNGQLDSNMNSSLNLSLSPHYGYGLMTPTWITTTITLNNNSAPTSATIAE 674
Oy 633 POKANR-----FHRLLTLTWLPAGYVPSKH 658
Db 675 QKTSIETFTPSNTTASIPNIKASGSGSANSNGEYITIKHTLVNMAVGYIVDPIR 734
Oy 659 RSPLIANTL---WGNMLATESLAKNSAEUTPSDHPFGITGGGGMVYODPRENHPGFH 715
Db 735 ROLLANSVLVHSGRNMTGLNSL-----PDNSWFLQGAATTLFTKQOKRLSYHGS 787
Oy 716 MRSSGYSAGMIA--GOTHTFSLKFSOTYTKLNERAKNNYSKNNY--SCGEMLFSLQE 770
Db 788 SASKGYTVSSQASGAHGHKFLFSQSSDKMKETKTNRLSRXYLSALCFEHPMD--- 844
Oy 771 GFLTKLVG--LYSGDNCHHFFYTOGENLTSOGTFRSOTMGANFPL-----PMKPFGS 824
Db 845 ---RIALIGAACNCGTHNMRSFY--GTRKSSKGRKFSHTTLASLRCLRDSMPR--- 895
Oy 825 THILAPFLGALGIYSSHFTVEGAVPRFSFTKPLINLVLPICGVKSGFMNATQORQAW 884
Db 896 -SMLTTPFOALFSTRERAPASIESDILARLFLBEQAHRAVVSPIQIKAYSSDTWPTLSW 954
Oy 885 TVELAYQPVLYROEGAIATOLLASKGIWFGSGSPSSRHAMSKYISQOTQPLSMTLTFLFOY 944
Db 955 EMELAYQPLLYKRPPLNTLLIQQNCSWVTNTPLAKHSF-YGRGSHSLKFSHLKLPANY 1013
Oy 945 HGFYSSSTFCNLYLNGEIALRF 965
Db 1014 QAEVATISTVSHYINAGALVF 1034
RESULT 9
G81722
polymorphic membrane protein E/F family TC0262 (imported) - Chlamydia muridarum (stra
C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: G81722
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heibelberg, J.F.; White, O.; Hicke
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
A:Reference number: AB1500; MUID:20150255
A:Accession: G81722
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1025 <TEF>
A:Cross-references: GB:AE002293; GB:AE002160; NID:97190298; PIDN:AAF39131.1; PID:9719
A:Experimental source: strain N199 (Mopn)
C:Genetics:
A:Gene: TC0262

Query Match 15.6%; Score 793; DB 2; Length 1025;
Best Local Similarity 27.0%; Pred. No. 7.5e-41;
Matches 269; Conservative 158; Mismatches 419; Indels 152; Gaps 35;

Oy 68 LOKTPMEGAAVTTIDYLSFFDTQKEGIFYFAKNLTPESGAGIYASPSPTVEIRDITIGPV 127

Mon May 7 09:06:32 2001

us-09-677-752-2.rai

Page 18

QY	519	ILKSGA	524
		:	:
		:	:
Dh	548	LTTTGA	553

RESULT 25
US-08-551-459-4

; Sequence 4, Application US/08551459
; Patent No. 5821350

GENERAL INFORMATION:

APPLICANT: Huang, Yue

APPLICANT: Karatzas, Costas N
APPLICANT: Lazaris-Karatzas, A

APPLICANT: Lazdalis, Annick
APPLICANT: Delaquis, Annick

TITLE OF INVENTION: ASPERGILLUS NIGER BETA GALACTOSIDASE

TITLE OF INVENTION: GENETICALLY MODIFIED BACTERIA FOR THE TREATMENT OF CANCER

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.
STREET: 335 Franklin Street

STREET: 225 Franklin Street
CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110-2

COMPUTER READABLE ;

MEDIUM TYPE: FL

COMPUTER: IBM PC compatible
OPERATING SYSTEM: DOS 3.31

OPERATING SYSTEM: PC-DOS
SOFTWARE: Patent In Belga

SOFTWARE: Patentlin Release #1.0
CURRENT APPLICATION DATA:

CONTRACT APPLICATION DATA:
APPLICATION NUMBER: US/08/551,4

FILING DATE: 11/11/2011

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: CJay
REGISTRATION:

REGISTRATION NUMBER :
PEEPENCE/DOCKET NIN

REFERENCE/DOC#1 NUMBER:
TELECOMMUNICATION INFORMATION

TELEPHONE: 617/542-5070

TELEFAX: 617/542-8906-

TELEX: 200154

INFORMATION FOR SEQ ID NO: 4

SEQUENCE CHARACTERISTICS:

LENGTH: 1007 amino acids
TYPE: amino acid

TYPE: amino acid
TOPOLOGY: 1 in

MOLECULE TYPE: protein

US-08-551-459-4

2000

Query Match	2.98;
Post Local Similarity	20.18;

Best Local Similarity	20.1%
Matches	200: Conservative 17

matches 200; conservative 1

Query Match	Score	DB 2;	Length
2.98;	146;	DB 2;	1007;
30.18;	Prod	NO	0.0014;

```

Db 250 GAVDIYGHDSYPIGIFDCANFTWMSG-----DLPTNRLHLBQSPPTPAIVEFG 300
Qy 312 AIYAPVTVLNDNGFTY-FINNIANKKGAIYIDGTSNKSISADRBHAIIFENIYTVNTNA 370
Db 302 GSYDPM-----GGGFPAACSELLNNEEVEFYKNDFFQIAINLNYIEGTMGNLGYR 356
Qy 371 NQST-----SANPRRNAIYVASSGCELLG--AGSSONLIFDPDIEVSAGVSVAFNKE 424
Db 357 NGYTSYDGSVYTESRN-TIREKISELKLGNRAKXSPGLTASFGCLTTSGL----- 408
Qy 425 ADQGSYVFSGATVNASADFHORNLOQTTPAPLTLSNGFLIEDHAAOLTVRFTQTVGVS 484
Db 409 ADTDLIY-----TPL-LGNSTQSFVYRHSDVSSSEST----- 442
Qy 485 LGNGAVLSCYNGANGSNASNAITLKIIIGLNLISLSCAEIPELWVEPNNNSNNTADT 544
Db 443 -----YKLRLPTSP--GSVYIIPOLG--GTLINGRDSKIHVDYVSGTINIYST 488
Qy 545 AATES-----LSYVKSLIDDDGNSVEESTDLTHALSOPMLSTSPASONOLRSDMDPSCI 601
Db 489 AEVFTWKFFADGKVLVL--YGGAG-EIHHL--AISTSNVYIEGSESGISSKOTSSVY 543
Qy 602 NVPIHGMGLTWCMWAKIQDEPPASSATITDPKXANRFRHRLTLTWLP-----AGVY 653
Db 544 -----VGMVYSTTRRIJOYGD-----LKLILDLRNSAYV-----WVPOLATQTSRPGFS 588
Qy 654 PSPKIHSPLIANTLWGNMLLATESLKNSAEITPSDH-----PEWGITGGGLGMVYODP 707
Db 589 TPEKVAASSIYK---AGLYVATYALKGSGLYLTADFNATTSVEYIGVSTAKMLFINGD- 644
Qy 708 RENHP-----GHHMSSSGISAGMIA-COHTHTSLKFQSOTYTKLNRY-----AK 750
Db 645 KTSHTVDKNGIWSATIVDYNAPDISLPSLKLDMKYVDTLPDIQSSYDSSLWPAADLKQTK 704
Qy 751 NNVSSKNYSOCEGMLFSLQEGFLTLTKLVLGS--YGDHNCH-----HFTYOGENLTSOGT 803
Db 705 NTLRS-----LTPITSLSKSDSDYCFHTGCLLYRGHFTATG-----NEST 742
Qy 804 FRSQTMGAVFEFDLPKPFQSTHLLTAPLGA-LGIYSSLSHTEFVGAVPRSESTKPLI 862
Db 743 FALDIOGSA-----FGSSVWLNGTYLGSMTGLYAN--SDYNAATYNLPOLOGAKTYVI 793
Qy 863 NVLYV-IGVKSFPNATROPAMVYELAYOVVLYROEGIGATQILASKJIMFGSGSSSR 921
Db 794 TVIVDINMLE-----ENMYTV-----GEDLKKTPRGILNIFLLA-----GRPSS- 830
Qy 922 IIMSYKIS-----QOTQPSLWTLNIOYHGFY 948
Db 831 -AISKULTGNLOGEDYEDKVGKPLNEGGLVAFROGCFH 866

```

Search completed: May 6, 2001, 19:17:29
Job time: 2144 sec


```

OY 608 -----WQGLMTGWAKTODPEPPASATITDPOKANRFRHRL----- 643
Db 545 GLEVRDNKIIVKLGANLRFENGAVTAGTVNPAPAPPTLTAEPPLRASNSHLQSLSEC 604
OY 644 -----LLTWLPAGYVPPKRRSP----- 662
Db 605 LVHNNALALQDGMENVQHGTLRVSGGLQMRGCIITVPPSGPIPIERPLAPLTOTEN 664
OY 663 -IANTLMGNMLLATESLK-----NSAELTPSDHPFWGITGGGLGMVYODPRENHPGFHR 717
Db 665 GIGLAGLELDESALQVKVGPGRMLNPEKRYVTLICPGIS-----FGQV-ANRTNYDVR 720
OY 718 SSGYSAGMIAGOT-----HTSLKFSQYTKLNERAKANNVSSKNSCGEMLFSLQ 769
Db 721 VS-VEPPVWFGQRGQTLFVLVGHGLHIONSKLQNLNQGGLKTDPTVNOLEVPVIGQGLEIAD 779
OY 770 EGFLLTKLVGLYSYGD-----HNCHEFTQGENLTSSQ-----TPRSOTMGCAVPE 815
Db 780 ESQVRKVL-----GDGLQFDSQARITTAAPNVTETLTGTSSMANVYRGTATAGSKLF 833
OY 816 DLPMPFGSTHILTAFLGALGIYSSLSHFTVEGAYPRSFSTKPLINVLVPIGVKGSFM 875
Db 834 -----LSLTFSTGLVGNMTIDSNAS-----FGQV-INGHEQIECFILLDQGNLK 880
OY 876 NATOPQAMTVE-----LAYOPVLYROE-----PGIATOLLASKGIMFSG----- 916
Db 881 EGSNLOGTWEVKNPNSASKAFLPSTALYPLINESRSLPG-----KNLYGMAILGGGCTCT 938
OY 917 ---SPSSRHMSYKISQO-----TOPLSWLTJLHFOY 944
Db 939 VIATLNGRRSNNPAGQSIIFWQEPNTIAROPLNHSTLTFST 981

RESULT 22
US-08-815-927-26
; Sequence 26, Application US/08815927
; Patent No. 6086890
; GENERAL INFORMATION:
; APPLICANT: MITTAL, SURESH K.
; APPLICANT: GRAHAM, FRANK L.
; APPLICANT: PREVIC, LUDVIG A.
; APPLICANT: BABIUK, LOURNE A.
; TITLE OF INVENTION: RECOMBINANT PROTEIN PRODUCTION IN BOVINE ADENOVIRUS EXPRESSION
; TITLE OF INVENTION: VECTOR SYSTEM
; FILE REFERENCE: 293102002101
; CURRENT APPLICATION NUMBER: US/08/815,927
; CURRENT FILING DATE: 1997-03-13
; EARLIER APPLICATION NUMBER: 08/164,294
; EARLIER FILING DATE: 1993-12-09
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 983
; TYPE: prt
; ORGANISM: Bovine adenovirus type 3
US-08-815-927-26

```

```

Query Match      2.9% Score 150; DB 3; Length 983;
Best Local Similarity 19.8% Pred No. 0.0061;
Matches 199; Conservative 108; Mismatches 392; Indels 304; Gaps 46;

```

```

OY 188 NTFVVENOSCFLEMDNICIQITNTAGKGAIVAGTSNFSFNCDLFFINACCAGCAIF 247
Db 37 NCFV--ENGEATLAM--LVEKPLTFDEKALTLGVRGIRINPAGLLETFND--LASAVP 89
OY 248 SPICSLTGRNGNIVFYNNCKFVETASSASDGAIAKVTTRLDVGNR-----GRI 239
Db 90 PLLAS--DEAGNVTJL-----NMSDGLYTKDKMLAVKVGPGSLSDSNNALOYHTDGLT 140
OY 300 FFDNDITKNYGAIVAPVVT-----LVNDCGPTYFINN--IANNK 336
Db 141 VTDDKVSIN-----TQAPLSTTSAGLSLLGLPSLHLEEEENLTIVNTAGLOISNNALAVKY 196

```

```

OY 337 GGATIIDGTSNKSIS-ADRIALIFENIV-----TN--VTNANGTSTSNPPRRNAI 385
Db 197 GSGITVDAQNQLAASLGGGLSRDKTKTVYKAGPGCLTITNOALITVATGGLQVNEGQIQL 256
OY 386 TVASSSG-----EILLAG-----SSQNLIEFYDPIEVSNAGVSFENKADQGS--- 430
Db 257 NITAGQGLNFANNSLAVELGSGLHFPQGNQVSLYP-----GDGIDIRNRTVYAGPGLR 312
OY 431 -----VVFSGATVNSADPFHORNLOTKTPAPLTLSNGF-----LCIEHQAULTVNR 476
Db 313 MLNHQLAVASG-----DGLVHSDTLRLKLSHGLTFENGAVRAKGLGTLTDDSGRSV--- 366
OY 477 TOTGVVSLGNCVAV--LSCYKNAGNSAS-----MASITLKHGLNLSLTKGAEI----- 526
Db 367 -RTGRGLRVANGQVOIFSGRGTAIGTDSLLTNIRAPLOFGPALTASLQSGPTIYNSN 425
OY 527 -----PLTWEPNTN-----NSNNYADTAATFSLSDVKLSLIDYGNSPY 566
Db 426 NCTFGLSIGPGMMVQONRLOVNPAGLVFGQNNLVPLNADPLAISDSKISLSLGPGLT-Q 484
OY 567 FSTDLTALIS-----SOPMLIS-----EASDNQLRSDMDPFGLVN-----PHYG- 607
Db 485 ASNALTLISLNGLEFSNQVAIAKAGRLRFESSQALESSILVNGTLTDTVIRPNLGD 544
OY 608 -----WQGLMTGWAKTODPEPPASATITDPOKANRFRHRL----- 643
Db 545 GLEVRDNKIIVKLGANLRFENGAVTAGTVNPAPAPPTLTAEPPLRASNSHLQSLSEC 604
OY 644 -----LLTWLPAGYVPPKRRSP----- 662
Db 605 LVHNNALALQDGMENVQHGTLRVSGGLQMRGCIITVPPSGPIPIERPLAPLTOTEN 664
OY 663 -IANTLMGNMLLATESLK-----NSAELTPSDHPFWGITGGGLGMVYODPRENHPGFHR 717
Db 665 GIGLAGLELDESALQVKVGPGRMLNPEKRYVTLICPGIS-----FGQV-ANRTNYDVR 720
OY 718 SSGYSAGMIAGOT-----HTSLKFSQYTKLNERAKANNVSSKNSCGEMLFSLQ 769
Db 721 VS-VEPPVWFGQRGQTLFVLVGHGLHIONSKLQNLNQGGLKTDPTVNOLEVPVIGQGLEIAD 779
OY 770 EGFLLTKLVGLYSYGD-----HNCHEFTQGENLTSSQ-----TPRSOTMGCAVPE 815
Db 780 ESQVRKVL-----GDGLQFDSQARITTAAPNVTETLTGTSSMANVYRGTATAGSKLF 833
OY 816 DLPMPFGSTHILTAFLGALGIYSSLSHFTVEGAYPRSFSTKPLINVLVPIGVKGSFM 875
Db 834 -----LSLTFSTGLVGNMTIDSNAS-----FGQV-INGHEQIECFILLDQGNLK 880
OY 876 NATOPQAMTVE-----LAYOPVLYROE-----PGIATOLLASKGIMFSG----- 916
Db 881 EGSNLOGTWEVKNPNSASKAFLPSTALYPLINESRSLPG-----KNLYGMAILGGGCTCT 938
OY 917 ---SPSSRHMSYKISQO-----TOPLSWLTJLHFOY 944
Db 939 VIATLNGRRSNNPAGQSIIFWQEPNTIAROPLNHSTLTFST 981

```

```

RESULT 23
US-08-614-377A-7
; Sequence 7, Application US/08614377A
; Patent No. 5976864
; GENERAL INFORMATION:
; APPLICANT: Smitt, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: No. 5976864ellini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF
; TITLE OF INVENTION: HETEROLOGOUS
; TITLE OF INVENTION: POLYPEPTIDES FROM CAULOBACTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson PC
; STREET: 225 Franklin Street

```


US-08-617-697-2

Query Match 3.5%; Score 176.5; DB 2; Length 1536;
Best Local Similarity 20.6%; Pred. No. 6.9e-06;
Matches 134; Conservative 98; Mismatches 236; Indels 181; Gaps 28;

QY 92 EGIYFAKULTEPESGACIGYASPNSTPVEIRDTIGYIENNITCC-----RPFY-139
Db 821 KGIYAKKNITEE-GGNITEGSKAVT-EIE---GNVTINNANVTLLIGSDPDHOKPLTI 875
QY 140 -----SSNMAAVNKRREGCAIHAONLYINHNHDVGFKNFYSYVGAISTANTFVY 192
Db 876 KKVYIYNSGLTAGNINVIAG-----NLVYSNANFKA-ITNFTFNVGGLFDKNGNSNI 929
QY 193 SENOSCFLEMD-----NICIOTNAGKGAIVAGTSNSESNNCDLFFINNACCA-----G 243
Db 930 SIAKGAFKPIDNSKNLSITNSSSTYRTIISG---NITKNGLDINTNEGSDTEMOIC 986
QY 244 GAIFSPICSLGNKGNIVFYNNRCK---NVTASSEASDGG-----AIKVTTRLDVT 293
Db 987 GDVSGKEGNLTISDKINITKOITIKAGVDGENDSDATINNANLTIKRELKLTODLNI 1046
QY 294 G-NRGRIFFSNITFNKVGAIYAPVTLVDNGPTVFNNIANNKGAIVIDGTSNSKISA 352
Db 1047 GPNKKEITAKDGSLLTNT-----NSADCTNAKKVTFNOKYKDSKISA 1089
QY 353 DRHAIIFENIYTVNANGTSTANP-----PRRMAITVASSGELL 396
Db 1090 DOKHTLHAKVETSGSNNTEDSSDNNAGLTIDAKNNVTNNNITSHKAKISATISGEITT 1149
QY 397 GAGSSONLI-----FYDPIKYSNAGVSVFKNKA---DQCSYVF-----SGA 436
Db 1150 KTGITINATTGNETVETAGTSTILGIESSGVTLTATEGALAVNISGNTVTANSGA 1209
QY 437 TVNSADFOHRIQTTPAPL-----TSLNGFLCIEDHAQTLVN-----474
Db 1210 LTTLAGSTIKTEGSEYTTSSQSDIGCTISGCTVEYKATESLTTOSNRKIKATTGEANVT 1269
QY 475 -----RFTGTGVVSLGNGAVLSCKYKNGAGNSASNAISTLK---HI-----512
Db 1270 ATGTCIGTSGNTVNAVAGDLTVGNCAGETINATECAATLTJSSGKITTEFASSHITSAKG 1329
QY 513 GINLSIILKSGAEIPLWEPNNSNNTADPAATPISDVYLSLID-----559
Db 1330 QVNLSA--ODGS-----VAGSINANAVTLNLTGT--LTTVGSNINMATSGTLVINAKDA 1379
QY 560 -----DYGNSPYESTDLTHALSOPMLISEASDQLRSDMDPSGLNV 603
Db 1380 ELNGALGN--HTVYVATNANAGSGVIATTSRVN--ITGDLITINGLNI 1425

RESULT 16
US-08-808-599A-24
Sequence 24, Application US/08808599A
Patent No. 6111089

GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Trophinin, Trophinin-Assisting
NUMBER OF SEQUENCES: 41
TITLE OF INVENTION: Proteins and Methods to Inhibit Implantation
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/808,599A
FILING DATE: 28-FEB-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/317,522
FILING DATE: 04-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/439,818
FILING DATE: 12-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 2256
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1160 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-808-599A-24

Query Match 3.4%; Score 174; DB 4; Length 1160;
Best Local Similarity 23.7%; Pred. No. 7.1e-06;
Matches 132; Conservative 62; Mismatches 200; Indels 164; Gaps 25;

QY 68 IOKTPNEGAATITDYLSFDTQKEGIYFAKULTEPESGACIGYASPNSTPVEIRDTIGPV 127
Db 544 LNSASFSGALSTSAFSGVNLGRAGFGGALNTNATFGCVLNGSGAFGAMNTNATPFGA 603
QY 128 IFENNTCCRPF-----TSSNPAAVNKRRE-GGAIHAONLYINHNHDVGFKNFESYVG 181
Db 604 LMSN-----AGFCGALISTFNGFGALNNSAGFCGAMNTSASFSGVLSAGF-----G 651
QY 182 GAIstantfVVSSENOscFLFMDNICIOTNTAGKGAIVAGTSNSESNNCDLFFINNACG 241
Db 652 GAINTSANFGGA-----LNSAGFGAI--STSAFSG-----ALNNSAG 689
QY 242 ACCGALFSPICSLTCNRCNIVFYNNRCKNVTASSFASDGAIAKVTTRLDVTGNRGRIF 301
Db 690 FGGAL-STSAFSGALNNSAGFGG-----AISTNASFGAI-----724
QY 302 SDNITKNYGAIVAPVTLVDNGPTVFNNIANNKGAIVIDGTSNSKIS-----351
Db 725 --SNPDPGGA-----STSVFCGTLNLTDFGSHNSNISISFSAPTTSVFCGSHSTNLC 778
QY 352 ---ADRHAIIF-----NENIV---TNTNANGTSTANPPRRNAIT---VASSSGEILLG 397
Db 779 FGGAPSTSLCFGSASNTNLCFGGSNTNCFSGATSANFEGSHISFGNGLSTAG---FG 835
QY 398 AGSSONLIIFYDPIEVS-----NAGVSVSNKE-AOQTG-----SVFSGATVSAADH 444
Db 836 NGLGTSAGDSSLGSTGCGSLGSPASFNGLGISTGCGGLGISTDPSGGLNHNADEN 895
QY 445 -----QRNLQTKTP-----APLTLNGLFLCIEDHAQTLV 473
Db 896 GGIAGNSAGFNGGLNINTDPCGELGTSAGCDLGSSTFGACGLVTSDF-----AG 946
QY 474 NRTOTGVVSLGNGAVLSCKYKNGAGNSASNAISTLKHLGLSLSLKSGAIPILLWEP 533
Db 947 NLGTINTGEGTLGTAGGFSVSLNN-GNGFGNPNASFNRLNTIIGFGG-----995
QY 534 TNNNNVTAD--TAATFS 549
Db 996 SNTSNGFTGPNTGSSFS 1013

RESULT 17
US-08-728-470-10

```

      821 KGIYAKKNITFE-GGNITFGSKRAVT-EIE-----GNVTINNANVTLLIGSDFDNHOKPLTI 875
      140 -----SSNPAVAVKIKREGGAIHAONLYINHHVDYGFMKFVSRYRGALSTANTFVY 192
      876 KKVIVINSNLTAGNIVNAG-----NLTVESNANFKA-ITNFTNVGGLFDNKNKSNLI 929
      193 SENOSCFLEMD-----NICIOTNAGKGAIVAGTSNFESENNCDLFFINNACCA-----G 243
      930 SIAGGARFKIDNSKNLSTTTSSSTYRTIISG---NITNKGDLITNEGSDTEMOIG 986
      244 GAIFSPICSLTGKRGIVYNNRCFK---NVTASSEASDGC-----AIKVTTLDDVT 293
      987 GGVSOKEGMLTISSDKINTKQITIKAGVDGENSDATNNANLITKTELKLTODLNI 1046
      294 G-NRGRIFFSDNTKKNYGAIVAPVTVLDNGPTVFINNINANKGAIYIDGTSNKSISA 352
      1047 GFKKAEITAKDGSDLTIGNT-----NSADGTNNAKVTENQYKDSKISA 1089
      353 DRHAIIFENIYNTVNTANGTSTSNP-----PRRAITVASSSGEILL 396
      1090 DGHKVTLSKSVETSGSNNTTSDSDNAGLITDAKNVTYNNNITSHKAVSISATSGEITT 1149
      397 GAGSSONLI-----FYDPIEVSNAGVSFENKKA-----DOTGSVVF-----SGA 436
      1150 KTGTTINATGNVEITAOGTGSLIGIESSGSVTLTATEGALVSNISGNTVTVTANSKA 1209
      437 TVNSADFHORNLQTKTPAPL-----TLSNGFLICEDHAOLTVN-----474
      1210 LTTLASTIKGTESVTTSSGSDIGTISGTYEVKATELTLTOSNKSIAKTGEANVTS 1269
      475 -----RFTGTGGVVSILGNAGVLSCKKNGAGNSASNASITLK---HI-----512
      1270 ATGTIGTISGNTVNTANGDLTVNGAEINATEGAATLTTSSGKLTEASSHITSAGK 1329
      513 GLWLSILKSGAEIPLWVEPTNNSNNTVADTAATFSLSDVKLSLID-----559
      1330 QVNLISA--QDGS-----VAGSINAANVTLLNTGT--LTTVKGSNINATSGTLIVINAKDA 1379
      560 -----DYGNSPYESTDLTHALSSOPMLISSEASDNOLRSDMDPFGSLNV 603
      1380 ELNGAALGN--HTVNVATNANGSGSVIATTSRVN-ITGDLITTINGLNI 1425

RESULT 13
US-08-469-880-2
Sequence 2, Application US/08469880
Patent No. 5876733
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Maltare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992

```

```

      PRIORITY APPLICATION DATA:
      APPLICATION NUMBER: US PCT/US93/02166
      FILING DATE: 16-MAR-1993
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 08/302,832
      FILING DATE: 16-SEP-1994
      ATTORNEY/AGENT INFORMATION:
      NAME: Berekteser, Jerry W
      REGISTRATION NUMBER: 22,651
      REFERENCE/DOCKET NUMBER: 1038-516 MIS-VG
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (703) 415-0810
      TELEFAX: (703) 415-0813
      INFORMATION FOR SEQ ID NO: 2:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 1536 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: DNA (genomic)
      US-08-469-880-2

Query Match      3.5%      Score 176.5      DB 2      Length 1536:
Best Local Similarity 20.6%      Pred. No. 6.9e-06:
Matches 134      Conservative 98      Mismatches 236      Indels 181      Gaps 28:

      92 EGIYFAKNITPFSGAGIYVSPSPVETIADTIGVIFENNNTCC-----RPFY- 139
      821 KGIYAKKNITFE-GGNITFGSKRAVT-EIE-----GNVTINNANVTLLIGSDFDNHOKPLTI 875
      140 -----SSNPAVAVKIKREGGAIHAONLYINHHVDYGFMKFVSRYRGALSTANTFVY 192
      876 KKVIVINSNLTAGNIVNAG-----NLTVESNANFKA-ITNFTNVGGLFDNKNKSNLI 929
      193 SENOSCFLEMD-----NICIOTNAGKGAIVAGTSNFESENNCDLFFINNACCA-----G 243
      930 SIAGGARFKIDNSKNLSTTTSSSTYRTIISG---NITNKGDLITNEGSDTEMOIG 986
      244 GAIFSPICSLTGKRGIVYNNRCFK---NVTASSEASDGC-----AIKVTTLDDVT 293
      987 GGVSOKEGMLTISSDKINTKQITIKAGVDGENSDATNNANLITKTELKLTODLNI 1046
      294 G-NRGRIFFSDNTKKNYGAIVAPVTVLDNGPTVFINNINANKGAIYIDGTSNKSISA 352
      1047 GFKKAEITAKDGSDLTIGNT-----NSADGTNNAKVTENQYKDSKISA 1089
      353 DRHAIIFENIYNTVNTANGTSTSNP-----PRRAITVASSSGEILL 396
      1090 DGHKVTLSKSVETSGSNNTTSDSDNAGLITDAKNVTYNNNITSHKAVSISATSGEITT 1149
      397 GAGSSONLI-----FYDPIEVSNAGVSFENKKA-----DOTGSVVF-----SGA 436
      1150 KTGTTINATGNVEITAOGTGSLIGIESSGSVTLTATEGALVSNISGNTVTVTANSKA 1209
      437 TVNSADFHORNLQTKTPAPL-----TLSNGFLICEDHAOLTVN-----474
      1210 LTTLASTIKGTESVTTSSGSDIGTISGTYEVKATELTLTOSNKSIAKTGEANVTS 1269
      475 -----RFTGTGGVVSILGNAGVLSCKKNGAGNSASNASITLK---HI-----512
      1270 ATGTIGTISGNTVNTANGDLTVNGAEINATEGAATLTTSSGKLTEASSHITSAGK 1329
      513 GLWLSILKSGAEIPLWVEPTNNSNNTVADTAATFSLSDVKLSLID-----559
      1330 QVNLISA--QDGS-----VAGSINAANVTLLNTGT--LTTVKGSNINATSGTLIVINAKDA 1379
      560 -----DYGNSPYESTDLTHALSSOPMLISSEASDNOLRSDMDPFGSLNV 603
      1380 ELNGAALGN--HTVNVATNANGSGSVIATTSRVN-ITGDLITTINGLNI 1425

RESULT 14

```

```
Db 648 TRKNTSYMTQSHDSHWNVSALNLETGANFTFKIYSSNSKGLTTOYRSSAGV----- 699
Oy 135 CRPTSSNPAVANKIREGGAIAHQ-----NEXTI 163
Db 700 -NFGVGNOMNSFN-LKEBAKVNFKLKPENNMNTSKPLRIRFLANTTATGCGSVFFDIYA 756
Oy 164 NHN-----HDVGFKNFNSYVRCAL-----STANTFVVSSEMSOCL 200
Db 757 NHSGGAELKMEINISNANFTLN-SHVRGDFAEKINKDLINATNSNFSLRQTKDFY 815
Oy 201 -FPMNICIOT-NTAGKGAIAAGTSNFSFNCDLFF-----INNA----- 239
Db 816 DQYARNAINSTINISLIGGNVTLGGONSSSTITGNTIEKANVTLEANNAPNOQNRDR 875
Oy 240 -CCAGGATFSPICSLTGN-----RCNIVFNNRCFK-----NVEASEASD 280
Db 876 VIKLSLVLNGLSLTGEADIKGNLTISESATFKGKTDTLNTGNTFNNGTAELNITQ 935
Oy 281 G-----GAIVVTRL-----DVTGNRGRIFSD----- 303
Db 936 GYVKLGNTNDDDLNITTHAKNRORSIIIGDIIINKKSLINTDSNDADIEIGIGNISQKE 995
Oy 304 -----NITK-----NYGAIYAPVTVLD----- 322
Db 996 GMLTSSDKINITKQITTKIGIDGEDSSDATSNANLTKTKELKLTEDLSIGFNKAEI 1055
Oy 333 -NGPTYFINNIANNKGA-----IYIDGTSNKSISADRHAIFFENIYTVNTNANGST 375
Db 1056 TAKDQDRLTIGNSDNGSCAEKATYTFNNVSKSKISADCHNTLNSKVATSSSGRESN 1115
Oy 376 SANPRRRAITYA-----SSSGEILLGASSONLIFYDPIEVSNAVY 417
Db 1116 SDND---TGLITTAKNVEYKNDITSLKTYNITASEKVITTAGST-----INATNGKA 1164
Oy 418 SVSPFNKEADQTSYVFCATVN-SADPHORNQOTKTPAPLTIJNSGFLCIEDHAQL----- 471
Db 1165 SIT-TKTGDISCT--SCNTVSATV---DLTKSGKIEKSGCANVTSATGTIGCTI 1218
Oy 472 -TVNRFOTGQVYSLGCAVLSCKYKNGASNASITLKHIGLISLKSAGIEPL 528
Db 1219 SCNTVFN-TYANAGDLTVGCAGIENATEGAATLTAIGNTLTTE-AG--SSITSTGOVUL 1273
Oy 529 L-----WVEPTNNSNNTYADTAATFSL---SDVKL---SLIDU-----YGNSPYSTDIT 572
Db 1274 LAQNSGSIAGSIANAANVTLNTTGTTLTVAGSDIKATSGTLVINAKDAKLMDASGDSDEV- 1332
Oy 573 HAISSOPMLISIEASDN--QLKSDMDQFSGLVN 603
Db 1333 NAVNASGSGSVTAATSSSVNITGDLNTVNGLNI 1365

RESULT 10
US-08-038-682-2
Sequence 2, Application US/08038682
GENERAL INFORMATION:
APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GENE III, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: Shoemaker and Matlare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/038,682
FILING DATE: 16-MAR-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERNSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-293
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single-
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-038-682-2
```

```
Query Match 3.5% Score 176.5 DB 1: Length 1536:
Best Local Similarity 20.6% Pred. No. 6.9e-06:
Matches 134; Conservative 98; Mismatches 236; Indels 181; Gaps 28;
```

```
Oy 92 ECIYFAKNITPEPGAGIYVSPNSPTVEIRDTIGPIFENNTCC-----RPT- 139
Db 821 KGIYAKKNITTFE-CGNITFGSKRAYT-EIE---GNVTINNANVTLLISDFPNHOKPLTI 875
Oy 140 -----SSNPAVANKIREGGAIAHQNLINHNHDVYGFKNFVSYRGCAISTANTFVY 192
Db 876 KRDVLIINSGLTAGCNIYVIAI-----NLTVESNANFKA-ITNFENNGLEDNNGNSMI 929
Oy 193 SNNOSCLFMD-----NICIOTNTAGKGAIAAGTSNFSFNCDLFFLINACCA-----G 243
Db 930 SJAKGARPKRIDNSKNISITTNSSYVRTIISG---NITNKGDLNITNEQSDTEMOIG 986
Oy 244 GAIFPICSLTGNRCNIVFYNNRCFK---NVEASEASDGC-----AIKVTTRLYVT 293
Db 987 GIVSKEGKLTJSSDKINTKQITTKIAGVDGNSDSDATNANLTKIKELKLTODLNI 1046
Oy 294 G-NKGRIFESDNIITKNYGAIAYPVTVLDNGPTYFINNIANNKGAIIYIDGTSNKSITA 352
Db 1047 GFNKAEITAKDQSDLTIGNT-----NSADGTNKKVTFQVKDSKISA 1089
Oy 353 DPHAIIFENIYTVNTNANGTSTANP-----PRRAITYASSGCEILL 396
Db 1090 DGHKVTLSHKVETSGSNNTEDSSDNAGLTIDAKNVYNNNITSHKAVISATSGEITTT 1149
Oy 397 GAGSSONLI-----FYDPIEVSNAVSVFNKA-----DQGSYVF-----SGA 436
Db 1150 KTGTTINATTCVETTTAOTGSLIGLIESSGCVTLPLATGALAVNISGNTVTVYANGSA 1209
Oy 437 TVNSADPHORNQOTKTPAPL-----TLNGLFLCIEDHAQLTVN----- 474
Db 1210 LTLTGSTIKTESVTTSSQSDIGITISGCTVEKATRESLTQSNKIKATGTGANTVS 1269
Oy 475 -----RFTQGTGCVYSLGCAVLSCKYKNGASNASITLKHIGLISLKSAGIEPL 512
Db 1270 ATGTIGGTTISGNTVNVTVNAGDLTVGCAGIENATEGAATLITSSGKLTTEASSHITSAG 1329
Oy 513 GLNLSIISKGAIEPLLWVEPTNNSNNTYADTAATFSLSDVKLSLID----- 559
Db 1330 QVNLISA--QDGS-----VAGSINANAANVTLNTTGT--LTVKGSININATSGTLVINAKDA 1379
Oy 560 -----DYGNSPYESTDLTHALSSOPMLISIEASDNQLKSDMDQFSGLVN 603
Db 1380 ELNGAALGN--HTVYNATNANGSGSVIATTSRVN-ITGDLITINGLNI 1425

RESULT 11
```

Db 1219 SCNTVN-VTANAGDLTVGNGCAEINATEGAATLTATGNTLITE-AG---SSITSTKGVDL 1273

Oy 529 L-----WVEPTNNSNNTYADTATFSL---SDVKL---SLIDD-----YGNSPYESTDILT 572

Db 1274 LAONGSIAGSIANAANVTLLTGTTLTVAGSDIKATSGTLVINAKDAKLGADSGDSTEV- 1332

Oy 573 HALSSOPMLISEASDN--QLRSDMDPFGCLNV 603

Db 1333 NAVNASGSGSVTAATSSSVNITGDLNTVNGCLNI 1365

RESULT 7

US-08-469-880-4

Sequence 4, Application US/08469880

Patent No. 5876733

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J.

TITLE OF INVENTION: High Molecular Weight Surface Proteins

TITLE OF INVENTION: Of No. 5876733-Typeable Haemophilus

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Maltare, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

STREET: Bldg. 1

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/469,880

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9205704.1

FILING DATE: 16-MAR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/US93/02166

FILING DATE: 16-MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/302,832

FILING DATE: 16-SEP-1994

ATTORNEY/AGENT INFORMATION:

NAME: Belkstresser, Jerry W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-516 MIS:V9

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1477 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-469-880-4

Query Match 3.5%; Score 177; DB 2; Length 1477;

Best Local Similarity 19.4%; Pred. No. 5.8e-06;

Matches 158; Conservative 115; Mismatches 244; Indels 296; Gaps 40;

Oy 37 KESISIKISLTD-----THNLITNCYLDNLRYIIALLLOKTRNEGAATVTTT- 82

Db 603 KDFRANVSLNLTGKGLNLTSSVNNLTNHLG-----TINISGNTINOT 647

Oy 83 ---YISFFTOKEGIFAFANLTPESG---AIGYASPNSP--TVEIRDTIGPIFENNNTC 134

Db 648 TKKNTSYMOTSHDSHMNVSALENTCANFTFIKIYISSNKGILITVOYRSSAGV----- 699

Oy 135 CRPTSSNPAANVKIREGCAIHAO-----NLXI 163

Db 700 --NENGVMGNMSEFN-LKEGAKVNEFKLPENNMNTSKPLPIFLANITATGGSYFEPIYA 756

Oy 164 NNN-----HDVVGFMKNFESYVRCGAI-----STANTFVSENGSCFL 200

Db 757 NMSGRAELKASEINISGANFTLN-SHVRGDDAFKINKDLITINATNSNSLRQTKODFY 815

Oy 201 --FMDNICIOT-NTAGKGAIACTSNSEFESNCDLEF-----INNA----- 239

Db 816 DGVARNAINSTYNTISILGANTVLLGGQNSSSTITGNITEKANANTLEANNAPNOQNTDR 875

Oy 240 -CCAGGAIFFPICSLTGN---RCNIVFYNNRCK-----WETASSEASD 280

Db 876 VIKIGSLVNGSLSLTGENADIKGNLTISESATFKGTRDTLNTGNETNGTAEINITQ 935

Oy 281 G-----GAIKVYTRL-----DVTGNGRGRIFPSD----- 303

Db 936 GVVKLGNTNDGDLNITTHAKRNORSIIIGDIIKKKGLNITDSNNDAEIOIGCNISQKE 995

Oy 304 -----NITK-----NYGAIYAPVTVYD----- 322

Db 996 GNLTISDUKINITKOITIKKGIDGEDSSDATSANLTIKTKELKLPEDLSISGPNKAEI 1055

Oy 323 --NCPYFINNIANNKGA-----IYIDGTSNKSISADRAHITFENAVTVNANGTST 375

Db 1056 TAKDGRDLTIGNSNDGSGAEAKVTFFNNVKDSKISADGHVYTLNLSKYKTISSNGREGSN 1115

Oy 376 SANPRRNAIYVA-----SSSGELLGAGSSONLIFYDPIEVSNGV 417

Db 1116 SDND---TGLITTAKNVEYKNDITSLKTVNITASKEVTTTGST-----INATNKA 1164

Oy 418 SVSEFKKADOTGSYVFSGATVN-SADFHQRNLQTPAPLTLNSGLCIEDHAOL----- 471

Db 1165 SJT-TKTGDISTGI--SCNTYVSATV---DLTTSKSGKIEAKSGEAVTATGTIGCTI 1218

Oy 472 ---YVNRFTOTGCVVSLONGAVLSCYKKGAGNSASNASITLKHIGLNLSSILKSAGAEPL 528

Db 1219 SCNTVN-VTANAGDLTVGNGCAEINATEGAATLTATGNTLITE-AG---SSITSTKGVDL 1273

Oy 529 L-----WVEPTNNSNNTYADTATFSL---SDVKL---SLIDD-----YGNSPYESTDILT 572

Db 1274 LAONGSIAGSIANAANVTLLTGTTLTVAGSDIKATSGTLVINAKDAKLGADSGDSTEV- 1332

Oy 573 HALSSOPMLISEASDN--QLRSDMDPFGCLNV 603

Db 1333 NAVNASGSGSVTAATSSSVNITGDLNTVNGCLNI 1365

RESULT 8

US-08-728-470-4

Sequence 4, Application US/08728470

Patent No. 5928651

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins

TITLE OF INVENTION: Of No. 5928651-Typeable Haemophilus

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Maltare, Ltd

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

STREET: Bldg. 1

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/038,682
FILING DATE: 16-MAR-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERKSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-293
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-038-682-4

Query Match 3.5%; Score 177; DB 1; Length 1477;

Best Local Similarity 19.4%; Pred. No. 5.8e-06;
Matches 158; Conservative 115; Mismatches 244; Indels 296; Gaps 40;

OY 37 KESLSNKRISLTGCD-----TNHLNLCYLDNLRILAIILOKTPNEGAAYTID- 82
DB 603 KDFRANNVSLNCTGKGLNISSVNNLTNLSG-----TINISGNITINOT 647
OY 83 ---YLSFDTOKEGIYFAKNLTPESG---AIGYASPNP--TVEIRDTIGPIFEENNTC 134
DB 648 TRKNTSYNOTSHDSHMVNSALNLETGANFTFIKYISNSKGLTTOYRSSAGV----- 699
OY 135 CRPTSSNPNNAVNKIREGCAIHAQ-----NLXI 163
DB 700 --NENGVGNMMSFN--LKEGAKVNEFLKPNENMNTSKPLDIRFLANTATGSGSVFEDIA 756
OY 164 NNN-----HDVGFPMKNFSYVVGAI-----STANTPVVSENOSEFL 200
DB 757 NNSGGAELKMEINISNCANFTLN--SHVGGDAFKINKDLTINATNSNFSLRQFKDFEY 815
OY 201 --FMDNICIOT--NTAGKGAITYAGTSNFSFNKDLFF-----INNA----- 239
DB 816 DGYARRNAINSTNISILGCONVTLGGONSSSITGNITIEKAANVTLEANNAPNOOIRDR 875
OY 240 --CCAGAIPIPSICSLTGN---RCNIVFYNNRCFK-----NVTASSEASD 280
DB 876 VKLGSLLVNGSLSLTGEVADIKGNLTISESATPFCKTDTLNTIGNFNGIAELINITO 935
OY 281 G-----GAIVYTRFL-----DYTGNGRRIFFSD----- 303
DB 936 GVVKLGNVTNODDLNITTAHRNORSIIIGGDI INKKGSLNITDSNDAEIQICGNISQKE 995
OY 304 -----NITK-----NYGAIYAPVTVYD----- 322
DB 996 GNLITSSDKINITIKOITIKKIGIDGEDSSDASNNALTIKTELKLTEDLSISGFNKAEL 1055
OY 323 ---NGPTFYNNANNGKA---IYIDGTSNKSISADRAHAIIFENINITYNTNANGSTST 375
DB 1056 TAKDGDLLTIGNSNDGNSGAEAKTVTFNNVYKDSKISADGHNVTLNSKVTSSNGGRESN 1115
OY 376 SANPPRRNAITVA-----SSSGEILLGAGSSONLIIFYDPIEVSNAGV 417
DB 1116 SDMD---TGULTITAKNVEVKNKIDITSLKTVNITASEKVTYTTAGST-----INATNGKA 1164
OY 418 SVSFNKEADOTGTSVSGATVN--SADFIQRNLOTKTPAPLTLNSGFLCIEDHQL----- 471

DB 1165 SIT--TKTGDISGTT--SGNTVSVSATV---DLTTKSGKIKENKSGEAVNTATGTIGCTI 1218
OY 472 ---TVNRFOTGTGVSLGNGAVLSCYKNGACNSASMTLKHIGLNLSTILKSGAIEPL 528
DB 1219 SGNTVY--VTANAGDITVNGAINEINTEGATLTATGNLTTE--AG---SSITSTKGOVDL 1273
OY 529 L---WVEPTNNSNNYADTATFSL---SDVKL---SLID-----YGNSPYESTDLT 572
DB 1274 LAQNGSINAGSINAANVTLTGTLTTVAGSDIKATSGTLVINADAKLNGDASGDSTEV- 1332
OY 573 HALSQPMLSTISESDN--QLRSDMDSEGLNV 603
DB 1333 NAYNAGSGSVTAATSSVNTIGDLTYNGLNI 1365

RESULT 5

US-08-302-832-4
Sequence 4, Application US/08302832
Patent No. 5603938

GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus

NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Matlare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

CITY: Arlington
STATE: Virginia

COUNTRY: U.S.A.
ZIP: 22202-0286

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,832

FILING DATE: 16-SEP-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1

FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US pct/us93/02166
FILING DATE: 16-MAR-1993

ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W

REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-404

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids

TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
US-08-302-832-4

Query Match 3.5%; Score 177; DB 1; Length 1477;

Best Local Similarity 19.4%; Pred. No. 5.8e-06;
Matches 158; Conservative 115; Mismatches 244; Indels 296; Gaps 40;

OY 37 KESLSNKRISLTGCD-----TNHLNLCYLDNLRILAIILOKTPNEGAAYTID- 82
DB 603 KDFRANNVSLNCTGKGLNISSVNNLTNLSG-----TINISGNITINOT 647
OY 83 ---YLSFDTOKEGIYFAKNLTPESG---AIGYASPNP--TVEIRDTIGPIFEENNTC 134

Mon May 7 09:06:32 2001

us-09-677-752-2.ra1

Page 2

Matches 190: Conservative 135; Mismatches 434; Indels 244; Gaps 36;

OY 41 SNKISLTGDT-----HNLNCTYIDNLRYLIALIOLKTPNGCAVTTIDVLSFPTQKEG 93
Db 602 SKTLLNADVAINELVYENNGSVOLNHNVTYLTITINAACQOILLVADPLNTTTLADG 661
OY 94 IYFAKNLPEEGGAIGVAPSPSYEIRDTIGPVFEENNTCCRPFTSSNPNAVAKIREG 153
Db 662 TNLGSAENPLSTIHFATYAAANDSI-LAVGKGVNLXANN-----ITTDANVAGSLHFRSG 715
OY 154 GAI-----HAQNLVINHNHDVYEFKNEFVSGAISTANTFVSNQSC-----198
Db 716 GTSIVSGVGGQGHKLNNLLDNGTIVKFLGDTTFNGGTRKEGKSILOISNNYTTDHYE 775
OY 199 -----FLPMNICTIOTNTAGCGAIYACTSNSPESNCDLFF--INNAACGCAIF 247
Db 776 SADNTGTLFVNTDPI---TTLNKGAYFGVLKQVYIISGPNIFENEGVNGIVHGIAA 832
OY 248 SPICSLTGRCNIVYNNRCFKNVETASEASDG-----GAIKYTRLDVYGNRGRI-- 299
Db 833 NSISFEMASLGSLFLPSTPLDVLITISTVNGIVDNFNAPIVVYSGIDSMINNGIIG 892
OY 300 -----FFSDN-ITKNYCGAIYAPVTVLVNPGPTPEINNIAKNGGAIYIDGTSNS- 348
Db 893 DKKNIIALSLGSDNITVN-ANTLYSGIRTKNNOGTITLGGMPNPGITVYGLGLENGS 951
OY 349 -KI-----SADRAHIIENENIVTNTNANGSTSNAPRRRAIVASSSGE- 393
Db 952 PLKQVTFETTDYNNLGSIIANNVYTIINDVTLTGCIAGTDFDAK-----ITLGSVNGNA 1005
OY 394 -----ILLGAGSSONLIEFYDPIEVSNAQ-----VSSEFNKEADQTSVYV 432
Db 1006 NVRFVDSTSDPSRMIVATQANKVTYVIGNALVSNIGSLDTPVASVRT--GNDGAGL 1063
OY 433 FSGATVNSADFHORNLOTKTPAPLTLNGLICIEDHAQLTVNRFOTGCVVSLGCAVVS 492
Db 1064 QGVIYQONIDFGTYNL-TILNSVILGGGTTAINGEIDLTNNLIFANGSTGMDTSSIS 1122
OY 493 CYKNAG-----NSASNASITLK-----HIGLINSLSLKSQAKIPLW 530
Db 1123 TTVLNVSSGNIGVYIAEDAOVNATTTGTTTICKIDNANANFSGTQAVYIIOGGRF- 1178
OY 531 VERTNNSNNTATATFSLSDVYKLSLID-----YGNSP 565
Db 1179 -----NGTLGAPRFAYTGSNIFVKYELIRDSNDYVLTRTNDVLANVTAVGSAIANAP 1233
OY 566 YESTDLTHALSOPMLISFASDNLSDMDFSGLNVPHYGMOGLMTWGMATQDDEPA 625
Db 1234 GVSQONISRCLES-----TNTAAYNNMLLAKPSPDVATFV-----GAITDTSAAV 1278
OY 626 SSAATITDPOKANRFHRTLLTWLPAGVYSPKRRSPLIANTLMCNMLLATESLKNSLELT 685
Db 1279 TTVNLNLTQKTOD-----LLSNRLCTLRVLSNAETSDVAGS---ATGAVSSGDEAE 1326
OY 686 PS-----DHPMGITGGGLGMVYODPRENIHGFHMRSSGSAGM--IAGOTHTFSLKESQ 739
Db 1327 VSTGVMAKPPYNI-----ABODKKGAGIAGYAKATTGVVYGGIDTLASDMLMGAICIGI 1378
OY 740 TATKL-NERYAKNNVSS-----KNYSCOGEMFLSLQECFLITKLVLGLXSY 783
Db 1379 TKTDIKHODYKKGDKDTINGLSFLYSQOLVKNEFAGNSIFTLNK-----V 1426
OY 784 GDHNCHEFTYOGEMLTQSO---GTFRSOTMGAVFF--DLPMKPGSGHTLILAPLGAIGI 838
Db 1427 KSKSORAFFEESNGMSKOIAAGNYDNMTPGCNLIFGYDYANMP-----NVJATPMAGISYL 1482
OY 839 YSSLIHTEVG---AYPSFSTKTPLLINVLVPIGVKGSFMMAT 878
Db 1483 KSSMENKETGTYYANKRINNSKFSDRVDLIYCAVAVAGSTVNT 1525

Sequence 9, Application US/08728470
Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728.470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Baretteser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ. ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1338 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-9

Query Match 3.5%; Score 177.5; DB 2; Length 1338;
Best local Similarity 21.4%; Pred. No. 4.5e-06;
Matches 154; Conservative 93; Mismatches 286; Indels 187; Gaps 33;

OY 75 GAAVTTIDVLSFPTQKEGIIYFAKNLPEEGGAIGVAPSPSYEIRDTIGPVFEENNTC 134
Db 546 GGNITIRQVEGTDSRVNKGVAANKNIT-FKGGNITFESOKA-TTEIK--GNVINKNTN 600
OY 135 CRPFTSSNPNAVAK-----IREGCAI--HAQNLVI-----NNHD 168
Db 601 A---TLGANTVAENKSPNLINAGVINNGNLTTAGSIINAGNLVSGANLOAITNTFN 657
OY 169 VVGFM-----KNESYVRGA-----ISTANTVVEENOSCFLEPMNICTIOTNTAGGAIY 219
Db 658 VAGSFIDNAGSINISIARGAKFKDINNTSSLNTITNSDTTY--RTIIGKMSKSG-- 711
OY 220 AGTSNFSNNDLFFINNACCA-----GGALFSPICSLTGRCNIVFVNNRCF-NVE-- 272
Db 712 -----DLNIIIDKSDAEIOIGNISQKEGNLTISDKVNTNTOITIKAGVEGG 759
OY 273 -TASSEASDGAIVVTRLDVYGNRGRIFFSDNITKNYGAIVAPVTVLVNPGPTPEYFINN 331
Db 760 RSDSEANANLTIOTKELKLAGD-----LNISGFNKAEIT--AKNGSDLTIGN 806


```

OY 590 QLRSDMDPSGLN-----PHYGQGLMTWMAK-TODPEPASATITDPQK 635
Db 585 -----FKGATVTKTGFPCDEIATPSHYGQKMSYTWMSRPLIPADCGPCGSPSPS 636
OY 636 ANRFRRTLLTW-----LPAGYVSPKHSRPLIANTLMGNMLLATESLKNSA-----EL 684
Db 637 AN---TLVAVWNSDTLVRSTYILDPERYGEIVSNSLW-----ISFLGNQAFSDILODV 686
OY 685 TPSDHPFMGITGGIGMMVYODPRENHGPFHMRSSSGYSAGMIACOTN--TFSLKFSQTYT 742
Db 687 LLIDHPGLSTITAKALGAVEHTPRGCHGFGSGRYGQYQALSMNYTDHTTLCLSFQLYG 746
OY 743 KLNERYAKNNVSSKNY--SCQEMLSLQEGFLTKLVGLYSYGDHNCHEFTYGENL-T 799
Db 747 KTNANPYDSRCSQEWYLLSFQGPPIVOKSEALISMKAAYGYSKNHNTTYLRPDKAPK 806
OY 800 SOGTFRSQTMGCAVFFDLPMKPFQSTHILFAP-----FLGALGIYSSLSHFTEVGA 850
Db 807 SOGQWHNNSYVLLISAE--HPFLNMCLLTRPLAOAMDLSGFISAEFLGQWOSKFTETGD 863
OY 851 YPRSFSTKPLINLVPIGVKGSFNNATOR--POANTVELAYOPVLYROEPGIATOLULASK 909
Db 864 LQRSFS-RGKGYNVSLPICSSQWFTPFKAPSTLTIKLAYKPDYRVNPNHIVTVVSNQ 922
OY 910 GIWFGSGSPSSRHMSYKISOOTOPLSMLTLHFQYHGFTSSSTFCNY 956
Db 923 ESTSISGANLRHGLFVOI-HDVVDLTEDTQAFILNYTFDGKNGFTNH 968

```

Search completed: May 6, 2001, 19:23:26
 Job time: 401 sec

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-J138;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hatlori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CWL029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brumham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA Bowman C., Dodson R., Gwinn M., Nelson W., Deboy R., Kolonay J.,
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [5]
RP Nucleic Acids Res. 28:1397-1406(2000).
RX EMBL: AE001628; AAD18591.1;
DR EMBL: AJ133034; CAB37069.1;
DR EMBL: AP002546; BAA98655.1;
DR EMBL: AE002192; AAF38163.1;
DR TIGR: CP0306;
KW Signal.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 928 OUTER MEMBRANE PROTEIN.
FT SEQUENCE 928 AA; 98332 MW; 58910A8F04F12219 CRC64;

Query_Match 11.5%; Score 585.5; DB 2; Length 928;
Best Local Similarity 24.9%; Pred. No. 11e-31;
Matches 26; Conservative 139; Mismatches 380; Indels 285; Gaps 46;

OY 1 MKKAEFFFLIGNSL-----SGLAREV-----PSRIFLMPNSVDP--TKESLSNKIS 45
DB 1 MMSLHMFLLISSIALPLSLNSAFRAVEINLGPITNSGCGTTPPQOTINADCTIYN 60
OY 46 LTGDTHNLTN-----CYLD---NLRYT---LAIQKTPNEGAAVTID----- 82
DB 61 LTGDV-SITNAGSPALTASCCKETTNLSFGHGQYQFLQNDAGANCTFTNTANKLL 119
OY 83 -----YLSFPTQKEGIFFAKNLPESGCAIGVSPNPTVEIRITQVIFENNTCR 136
DB 120 SFGSGSYLSLIOT-----TNATTTGA-----INSTGACSTIQSNVSCYF 158
OY 137 PFTSSNPNAVVKIREGAIHAQNLVYNNHVDVGFKNFSYVVGGAISTANTFVYSEHQ 196
DB 159 GGNFNSNDN-----GGLQSSSISLSLNPLT--FAKKKATOKGALXSTGITTNNLT 209
OY 197 SCFLFMONICITNTAGKGAIACTNSFESNCDLFTINNACCA-----GGAISPTCS 252
DB 210 NSASFSSEN-----TAANNCGAIYTEAS--SFISNKAIISFINNSVATSAITGAIY--CS 260
OY 253 LTGNNGNIVFYNNRCFKVETASSEASDGAIKVTRLDVLTNRRGIFESDNTKNYGA 312
DB 261 STS-----APKPVLTLS-----DNGNLNIGNTAITSGGA 290
OY 313 IYAPVVTLVNCPPTFYINNANNK-----GGAITYIDGTSNKSISADPHAIFFENITVNT 368
DB 291 IYTDNLVLSGSPTLFKNNSAIDTAAPLGAIAIDSGSLSLALGGDITFEGRNTVYVKA 350
OY 369 NANGSTSNAPRRNAIYVASSGELL-LGASSONLITYDPEVLS-NAGVSYSEFKKED 426
DB 351 SSSQITTT-----RNSINIGNTNAKIVQIRASOGNTIYDPTITTSITALS DALNLNCP 404
OY 427 Q-----TGSVVFSGATYNSADPHOR-NLOTKTPAPLTLNSGFLICEIDHQAOLTVRFTQ 478
DB 405 DLACGNPAYOGTIVTFSGEKISEAEADNLKSTIOQPLTLAGGQSLKAGCVTLVAKSFSQ 464
OY 479 TGGVVS LGNVAVLSCKYKNGNSASNA-SITLKHIGLNLSTILKSGAEIPLLVETPNNS 537

DB 465 SPGSLTLM-----AGTTLETADGITTNNLVANDSL-----KETKKA 502
OY 538 NNVTADTATFESLDVKKSLIDDYG-----NSPYESTDLTHALSSQPMISEASDN 589
DB 503 TLKATQASQTVTLSC-SLSTLDPGSGNVYEDVSMNNQVFSCLTLADDPANHITDLAAD 561
OY 590 QLRSDMDFSGLNVPHYXGQGLMTQCMATQOPEPARSSATITIDPOKANFHHTLLTMLP 649
DB 562 PLEK-----NPIHWGTOGNMALSW---QEDTATKSKAAT-----LTWTK 597
OY 650 AGVYSPKHSPLANTLILGNNL-----LATESLKNASLETPSDHPWCITGGGLGM 702
DB 598 TQYNPNPERGTLVANTLMGSEVDVRSIQLVATVYRQGEIR-----GIMCGISIF 650
OY 703 VYQDPRENHGFHMSSGYSAGMIAGQHT-----FSLKSQTYTKLNERYA-KNNVSS 755
DB 651 FPKDSKTKINKGFRHISAGY-----VVCATTTLASDNLITAAFCQLPGKDRDHFINKNRASA 706
OY 756 KNYSGOGEMLFSLQGFILTKLVG-----LTSYGDHNCHEHTYO--GEN-- 797
DB 707 YAASLHLQHLATLSSPSLURLPGSESEQPVLPDAQISYISKNTMKTYTQAPRGESSW 766
OY 798 -----LTSQGTFRSQTGCAVFPDLPKPGSHIILTAPELGAIGYS 840
DB 767 YNDGCALELASSPLHTALSHGELFHA-----YF-----PFIKVEASTI 804
OY 841 SLSPHTENG-AVPRSPSTKPLINLVLPVIGKGSFPMATQRPQAVTVELAYQPVLYROEP 899
DB 805 HODSFKERNTTLVRSFDS-GDLINVSVPICITFEFRSNER-ASYEATVIYADVYRRNP 862
OY 900 GIATOLLASKGIWPGSGSSSRHA-----MSKYSQOTQPSLTLHFQ 943
DB 863 DCTTALLINNTSKTGTGNLSRQAGIGRAGIGYAFSPNLEVTNLSMEIR 912

RESULT 24
ID 092895 PRELIMINARY: PRT: 978 AA.
AC 092895;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN H FAMILY.
GN PMP-14.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydiophila.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CWL029;
RX MEDLINE=9920606; PubMed=10192388;
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."
RL Nat. Genet. 21:385-389(1999).
DR EMBL: AE001628; AAD18596.1;
DR INTERPRO: IPR000515;
DR PROSITE: PS00402; BPD-TRANSP-INN_MEMBER; UNKNOWN_1.
SQ SEQUENCE 978 AA; 103655 MW; 12D6610CBEB0871 CRC64;

Query_Match 11.5%; Score 583; DB 2; Length 978;
Best Local Similarity 26.1%; Pred. No. 1.8e-31;
Matches 263; Conservative 137; Mismatches 411; Indels 196; Gaps 45;
OY 41 SNKISLTGDTNHL--TNCYL-DNLRVYLAIILOKTPNEGAANTIDYLSFFDTQKEGIYFA 97
DB 67 SGNLNLGLGSLTFTSCAPRTNSNALL-----SAAETLT-FKNF-----SSINFT 112
OY 98 KULTPESGCAIGVAPSPFVEIRDTJGVIFFENN-----TCRPTTSSNPNAVVKIREG 153
DB 113 GNSGTGLGGLI-YGRD-----IVFOSIKDLIFTIRVAVSPASVTSATP--AITTVTTG 164

084880
ID 084880 PRELIMINARY: PRT: 1016 AA.
AC 084880:
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE PUTATIVE OUTER MEMBRANE PROTEIN H.
CN PMPH.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_Taxid=813:
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/CX:
RC MEDLINE=99000809; PubMed=9784136;
RX Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.:
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
DR EMBL: AE001360; AAC68470.1;
DR INTERPRO: IPR002016;
DR PROSITE: PS00435; PEROXIDASE_1; UNKNOWN_1;
SQ SEQUENCE 1016 AA; 107904 MW; E691912C3A2BDE67 CRC64;

Query Match 11.7%: Score 595; DB 2; Length 1016;
Best Local Similarity 25.4%; Pred. No. 2.9e-32;
Matches 261; Conservative 149; Mismatches 404; Indels 214; Gaps 49;

QY 42 NKISLTGDTHTNLT-----NCYLDNLRYLILQKTPNEGAAYITDY, SFFDTQKEGIYF 96
DB 69 DMLTTIGQNHHTLSFTDSCGPVLONTAFISA-----GETLTKPFSS-----LMF 112
QY 97 AKNLTPESGAGICYASPNSTFEIRDTIGPVIFENNT-----CCRP 137
DB 113 SKNVSGCEKGM-----SCKTVSISGA-GEVJFMDNSVGSPLSIVPASTPPAPAPAP 166
QY 138 FTSSNPNAVNKIREGGAITHQNLINHNHDVYGFKNSTYVAGAISTANTVSENGS 197
DB 167 AASSSLSPYSDARKSIFSVET-----SLGISGVKGVNF-----DNNAG 207
QY 198 CF--LFMDNCTIOTNTAGKAGIYAGTNSFEENCD--LFFINNACAGAIFFPICS 252
DB 208 NFGTVRGN---SNNNGSGS-GSATTPSFYKCKGKXSFIDNVAACGGGVYKGTVL 263
QY 253 LTCNRCNIVFYNNRCRNKVTASSEASD-----GGAI-KVYTRLDVTGNKGRIFESD 303
DB 264 FKNNEGIFFRGNTAYDGLITLAATSHDQNTETGCGGCVICSPDDSVKPEGNGSIVFDY 323
QY 304 NTRKNTGGAIAVAVTLVDNGPTFYFININNGGAIY---IDGTSN-----KISAD 353
DB 324 NFAKRGCSILTKREFSLVADDSVAFSNNTEKKGGAIAVPTIDISTGSGILTEBRNRAE 383
QY 354 RHAIIFNE-----NIVVTNTA---NCTSTANPPRRNATIVASSGEBILLAGSSO 402
DB 384 GGAICVSEASSTGNTLTLSDDDIYFSGNMTSDRGERSAAILLDGTTVSLNAGSL 443
QY 403 NLJFYDIEVSN--AGVS-----VSFNKEADQIGSVFSCATVNSADPHQ--RN 447
DB 444 KLIFDYDVVQNSAAGASTPSSSSMPGAVTINOSGN--GSVIFTAESLTPSEKLDVYN 501
QY 448 LOTTPAPILTSNGFLIEDHAQILTVRFQGTGVSIGNAGVSCYKNGANGASNASI 507
DB 502 STSNPPALTYVSGELVATVTCATITTCGTITATSGRVTLGASASIAVAGANN---NTC 558
QY 508 TLKHTIGLNLSSILKSGAIEPIIAWEPTNNSNNYADTAATFESLDVYKLSLIDYGNSE 567
DB 559 TVSKLIGIDLEFLTPNKITAILGADG-----VTVNSGSTLDL--VMESEAEVYDN-PLF 610
QY 568 STDLTHALLSOPMISISEAS-DNQLRSDDMPFSGIINPHTYMOGLMTWGMKATQDDPEAS 626

DB 611 VGSLLT-----IPVTLSSSSASNGVTKNSVTINDADAHYGCGSSADMTK----PPLAP 662
QY 627 SATITDPOKANRFRHTLLTLWLPAG-----YVSPKHSRSLPIANTLW--GNMLA-TESLK 679
DB 663 DAKGVPPENTN---NTLTLWRPASNVGEYRLDPQRKELVPSNLMVAGSALRTFTNGLK 719
QY 680 -----NSAELTPSDHPHFGITGGCIGMMVYODPRENHGCFHMRSSGYSAGMIAQOTH--- 731
DB 720 EHYVSRDVGFAVSLH-----ALGDYILNVTODRDRGFLARYGFOA---TASHVEN 768
QY 732 --TESLKFSQRY--TKLNBRYAKNNVSSKNYSCGEMFLSLOEGFLTKVLGY--SYGD 785
DB 769 GSIFGVAFGQLYGOTSKRMNYSKDGNNMTLSCFGRSVVDIKG---TETVWVETATG- 823
QY 786 HNCHEIFTYOGENTLSQGTFRSQ-----TMGA---VFEDLPMPFGSTHILTA-- 830
DB 824 YSVHMHHTOYFNDKTQKRDHSCWHNNNNYAFVGAENHFLCYCIPTRQFARDYELTGFM 883
QY 831 PFLGALGIYSSLSHFTFVGANPRSFSTKTPILNVLVPIGVKSGFNNATQR--POAWYEL 888
DB 884 RFEMAGWSSST--RETGSLTRYFARSG--HMSLPIGIYAAVAVSHVRSPPSKLTLM 939
QY 889 AYOPVLYRDEPIATOLLASKGIMFGSGSPSRHMSKXISQOTOPLSWLTJLHFQHGFI 948
DB 940 GYRPDIWKVTPHCNMELIANGVKTPIQGSPLARHAFLEVDH-----TLXIHHC-- 989
QY 949 SSSTFCNY 956
DB 990 --RAYMNY 995

RESULT 19
QY 09JSE2 PRELIMINARY: PRT: 947 AA.
ID 09JSE2:
AC 09JSE2:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE POLYMORPHIC MEMBRANE PROTEIN A FAMILY.
GN PMP-19.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_Taxid=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hiraoka H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishi K., Hattori M., Kihara S., Nakazawa T.:
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL: AF002547; BAA98745.1;
SQ SEQUENCE 947 AA; 103628 MW; 9CBFDAF290A771EC CRC64;

Query Match 11.5%: Score 587; DB 2; Length 947;
Best Local Similarity 25.2%; Pred. No. 9.2e-32;
Matches 250; Conservative 174; Mismatches 389; Indels 178; Gaps 47;
QY 14 LSGIAREVPSRIFLIPNSVDPDPKESISNNI---SLTGDTHNLITNCYLDNLRYLALILOK 70
DB 30 LSGHSGEDLEFLTRSSSPKTTIYSLRKDFIYCDFGNSIHKPGAFLWLGKDLFFINS 89
QY 71 TPNEGAVTTIDYLSFDTQKEGIYFAKNLTPESGAGIYASPNSTFE-----IRDT 123
DB 90 TPL--AAITFRKI--HIGAGAGAGLFSGSNT-----FKGLSLVLENNESGCVLTT 137
QY 124 IGVPIFENNT--CCRPSTSNPNAVNKIREGGAIAHQN-----LYNHNHDVYGFKNF 176
DB 138 SGDSLFINNTSVLCQNNISYCP-----GGALLQGRKSKALFFRDRNGTILFLKNK 188
QY 177 SYVR-----GCAISRTANFVSENGSCFLFMN--ICITNTAGGAIYACT--SNSF 226

```

OY 104 SGAIGYASPNSTVEIRDT---IGPVIFENNTCCRPFTSSNPNAVKIREGAIHMO 159
DB 103 AOSNMAAATAADKALYLTGFSNLSFIAPAGTTVASGSLSSAGALINTDNGTLFSSQ 162
OY 160 NLYINNHVDVGFEMKNFYSVGAIASTANTFVSENOGFLFMDNICIOTNTAGK-GGAI 218
DB 163 NYSNENNNM-----GGAIIT-KTLSISGNTS-----STFTPSNKAKTGGAI 203
OY 219 YACTSNFESNNCDLFFINN-ACCACGAI-FSPICSLTGNGNIYVYNNRCEKNVETASS 276
DB 204 YSSAASISGNTGOLVFEMNNKGETGGALGFEMASSITON-SSLFSGN-----TAD 255
OY 277 EASDGAIV-----TTRLDVGNRGRIFESDITKNYGAIAVPVYVDNGPTYEINN 331
DB 256 AAKGGAITCEKGETPTLTISGNS-LTFAENSSVYOGAICAHGLDIASAAPTLPFSA 314
OY 332 IANN-----KGAIYIDGTSNSKISADRAHIIENENIVTNTANSTSTSNPRRAITV 387
DB 315 RCGNTAAGGGAIAIADSGSLSSANOGDITLGNLT-----TSTAPSTRNAIYL 366
OY 388 ASSSGEITLGAASSONLIFYDPIEVSNAGVS-----VSFNKADQTSVPSGATVNS 440
DB 367 GSSAKITNRAAOGOSIYFDPASNTGASDYLITNOPDSNPDLPSGTYFSGEKL-S 425
OY 441 ADFHOR--NLQTKTPAPLTLNGLFCLIEDHQAULTVNRFTQTVGVSLSGNAVLSCYKNGA 498
DB 426 ADEKADADFTSLKQPLASGTLAKGNVELDVNGFOTEGSTLL-----NQPT 477
OY 499 GNSASNAITLKHIGLNLISLKSAGELPLWEPNNNSNYADTAATESLSDVKLSLI 558
DB 478 KXKADFEALSLTKLVYDDEL-----EGNKSYSIETAGANKTITLTS-PLVQ 524
OY 559 DDYGNSPYSTDLTHALSSQPMLSISEASDNQSRDMDPSGLNV-----PHYMOGL 611
DB 525 DSSGNF-YESHITNQAF-TQPLVVFYAT--AASDIYDALTLSPVOTPREPHYGOGH 578
OY 612 WTGWMKATDDEPFASSATITDPKANRPHRTLLTLWLPAGVYSPKHSPLINTLWGN 671
DB 579 WEATMA--DTSIAKSGT-----MTWVTGYNPNPEKRSVVDSDSLWAS 620
OY 672 LATESLKSASALTPSDHPFW---GITGGGLGMVYQDPRENNHGFHMRSSGYSAGMIAG 728
DB 621 --TDIRTIQOIMTQANSIYQORGLMAGTANFHKDKSGTQARRHKSUYGIVGSAE 677
OY 729 --OTHTFSLKFSOTYTKLNERIAKNNVSKNSCO-----GEMLPFSLQ 769
DB 678 DESENIFSAFCOLFCKDKDLFIVEN-TSHNYLASLYLQHRAFLGGLPMPSEFSITDMK 736
OY 770 EGFLLKLKLVGLSYGDNCHNFT-----OGENLTSOGFRSOTMGGAFFDPMK-PRG 823
DB 737 DPLILNAOLSTSYTKNDMDRTTSTPEAOGSVTNNSGALE---LGSIALYLPKKAPEF 793
OY 824 SHILTAPELGAIGYSLSHFTVEGAYPRSESTKPLINVLVPIGVKGFMAATORPA 883
DB 794 OGF---PFLKQAYVSRQONFESGAEARAFD-DGDIVNCSTPVGIRLEKISEDEK-NW 848
OY 884 WVELAYQVLYRQEPGIAITOLIAKGIW 912
DB 849 FEISLAVYIDGVYRKNRSTRSLWASGW 877

```

RESULT 16

```

OY 086164 PRELIMINARY: PRT: 928 AA.
AC 086164:
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, last annotation update)
DE OUTER MEMBRANE PROTEIN 4 PRECURSOR.
GN OMP4 OR PMP-11.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.

```

```

OX NCBI_TaxID=83558:
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD/CWL-029/YR-1310;
RA Knudsen K., Madsen A.S., Mygind P., Christiansen G., Birkelund S.;
RT "Identification of two novel genes encoding outer membrane complex
associated surface layer proteins in Chlamydia pneumoniae."
RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VR1310;
RA Hjerno K., Boesen T., Dagaard L., Knudsen K., Madsen A.,
RT Christiansen G., Birkelund S.;
RT "Chlamydia proteins containing the GGA1-repeat belong to a subfamily
of autotransporting pathogenicity factors."
RL Submitted (Jan-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lamell C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."
RL Nat. Genet. 21:385-389(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=CJ38;
RX MEDLINE=2030349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CWL029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL: AJ001311; CA04672.1;
DR EMBL: AJ13034; CAB37072.1;
DR EMBL: AE001628; AAD18593.1;
DR EMBL: AP002546; BAA98658.1;
KW Signal.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 928 OUTER MEMBRANE PROTEIN 4.
SQ SEQUENCE 928 AA: 98903 MW: 788BCDD62C911402 CRC64:

```

Query Match 11.7%: Score 597; DB 2; Length 928;

Best Local Similarity 26.1%: Pred. No. 1.8e-32;

Matches 253; Conservative 148; Mismatches 361; Indels 206; Gaps 47;

```

OY 36 TRESLSNKISLTGD-----THNLTYCY--LDNLRYI-----LAIQKTPNEGA 76
DB 45 TPKTSATITSLTGDVFFYEPKGTPLSDSCFQKOTTNLFLGHSLTGCFIDAGHAGA 104
OY 77 AVITIDYLSFPDQKEGIFYFAKNLTPESGAGIYASPNSTVEIRDTGIPVIFENNTCCR 136
DB 105 AASTT-----ANKNLTFSGFSLSFSSPSTTV-----TTGQ----- 136
OY 137 -----GTLSSAGAVNLENI--RKLVAAG--NESTADGGAIGA-SFLTGTGS 178
DB 197 SCLEFMDNICIOTNTAGKGAIVAGTSNFSFNCDL--FFINNACCAGAGATSPICSLT 254
DB 179 GDALEFSNN-----SSSTKGAIT-ATTAGARINANTGYVFLSLIASTSGAIDDETSIL 232
OY 255 GNRGNIVFYNNRCFKNVEFASSEDGCAIKVTRLD---VTGNRGRIFESDITKNYG 310
DB 233 SNKKFLYFEGN-----AAKITGAIICNTKASGPELLISNNKTLIFASNAETSG 282
OY 311 GAIYAVVTVLVNDGPTYFI-NNI--ANNKGAIIYIDGTSNSKISADRAHIIENENIVTV 367
DB 283 GAIHAKKALSSGGLTEFLRNNVSSATPKGATSIDASELSIATGCTITVRRNTLT-- 340
OY 368 TNANGTSTSNPRRAITVASSSGEITLGAASSONLIFYDPI--EVSAGVSVSFNKEA 425

```

```

Db 519 TCKEYENKLANDTLA---LGGIOLG-----GKSVTTTNVSHV----- 555
Oy 591 LRSDMDFSGLVNPHYGMGLWTMGACTODEPASPATITDPOKANRPHRTLLTLPLA 650
Db 556 -----GVAETHYGVGNMVSVMKDNNSDPKT-----QTAIFTMNKT 592
Oy 651 GVPSPKRRSPLIANTLMGNMLLATESLKNSAELTPSD--HPPWGITGGGLCMNVYODR 708
Db 593 GVPNPERAPLVLNLSLMS--FIDLRISQDVLERSYDSILETRGLMVGSGIFHFKDRN 651
Oy 709 EHHPPGHMSSSGYSAGMIGOTH--TFSLKFSQTYTKLBERAKNNVSS----- 755
Db 652 AENRFRHISISGVYGLATTNTSREDLSVAFQOLFAKDDYLVSKAAVAVGASYOYHV 711
Oy 756 -----KNYSCGEMFLSLOEGFLTKLVGLYSYGDHNCHEFTOGENLTSQGT 803
Db 712 SKFDLITLRFNPNPTCCSG---FSKEIRIFLDQAQI--TYCHTANNMTSTYDPEV--KGS 765
Oy 804 PRSQMGAVFEDLPKPPGSTHILIT--APFLCALGIYSSLSHTEVAVPRSPSTKPLI 862
Db 766 WONDITLGLTSLTSPDIPVSSSIFDSYAPFAKQVYVAHQDDFKETTEGKRVFES--SDL 824
Oy 863 NVLVPVGVSGFMNATFORQAMTVELAYQVLYROBPGLATOLLASKGIMFGSGSSRH 922
Db 825 NVSPVPIGFEKLYGER--SAYDLTLMYLPDYVRHNPSCMTGLAINDVSWLTTATNLARQ 883
Oy 923 AMSYKISQOTPLSMTLHFQYHGFYSSSTFCNY---LNGEIA 962
Db 884 AIVRAGNHIALTSGVEMFSQF--GFELRSSSRNYNVDLGAHY 925

RESULT 13
O9PL45 PRELIMINARY: PRT: 987 AA.
ID 09PL45:
AC 09PL45: 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DE POLYMORPHIC MEMBRANE PROTEIN G FAMILY.
GN TC0263.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / NIGG;
RX MEDLINE=20150255: PubMed=10684935;
RA Read T.D., Brumham R.C., Shen C., Gill S.R., Heidelberg J.F.,
  White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,
  Linher K., Weidman J., Khouli H., Craven B., Bowman C., Dodson R.,
  Gwinn M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
  Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
  pneumoniae AR39."
RT Nucleic Acids Res. 28:1397-1406(2000).
RN 12
RP SEQUENCE FROM N.A.
RA Read T.D., Brumham R., Shen C., Gill S.R., Heidelberg J.F., White O.,
  Hickey E.K., Peterson J., Umayam L.A., Uterback T., Berry K.,
  Bass S., Linher K., Weidman J., Khouli H., Craven B., Bowman C.,
  Dodson R., Gwinn M., Nelson W., Deboy R., Kolonay J., McClarty G.,
  Salzberg S.L., Eisen J., Fraser C.M.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AEO02293; AAF39132.1;
DR TIGR: TC0263;
SQ SEQUENCE 987 AA: 104667 MW: 32079866BEBB2UA42 CRC64.

```

```

Query Match 12.2% Score 620.5; DB 2: Length 987;
Best Local Similarity 25.8% Pred. No. 5e-34;
Matches 253; Conservative 153; Mismatches 404; Indels 171; Gaps 42;
Oy 39 SLSNKSISLTGDTNLTNCTYLDNLTALITLAIQKTPNEGAA-----VTTDTVLSFF 87

```

```

Db 82 NLGNFTTIGRSHSLV---FENIR-----TSTNGALSNHABSGFLVIEAFDELSSL 130
Oy 88 DTQKEGIVFAKULTPESGGAIGVAPNSPTV-----EIRDTIGPIFENNTCCRPFTS 141
Db 131 NCNS-----LVSVFOTGTT--TSVPSNCTIYSRDLVLRD--IKKVSFSLVS----- 177
Oy 142 NPNAAVNKIREGAIHAONLYINHNDVYFMKNFSYVKGAI1STANTFVSENSCFLF 201
Db 178 -----GDCGALDAOGLMNVNGIEKLTCTFOENVAQSGCAGCYQTKTPSAVGNKVPISF 228
Oy 202 MDNICIQTTAPKGAIIYA-----GTSNPF-----SNMCDLPIINNACCAGALFSPIC 251
Db 229 LGNV-----AGNKGCGVAAVKDGCGAGATDLSVNEFANTVAVEFENSRARIGGIIYSD-- 281
Oy 252 SLTGNRGIVFVNNR---CFKNVETA-----SSEASDGAII--KVTTRLD 291
Db 282 -----GNISFLGNKATVFLSNVNSPIIYVDPAAAGQPPADKDNVGDGALFCKNDNTIG 335
Oy 292 VTG--NRGRIFESDNTIKXNYGAIYAVVTLVNDGPFPINNIAANKGAIYIDGTSNK 349
Db 336 EVSFKDEGVVFEFSKNIAAGKGAIIYAKKLIISDCGPVQFLGVAND--GCAIYLVQDEL 394
Oy 350 ISADRAIIFENIYTNVNTANCTSTANPPRRNATTVASSGEILLGSSQNLFIYTP 409
Db 395 LSADRGDIIFDGNLKRMAIQAATYADV--MVASNAISMATGQIITTLRAKEGRRIIFNDP 453
Oy 410 IEVSNAG---VSVPKKEADQTSVYES--GATVNSADFHQRNLQRTKPAFLTNSGFLCI 465
Db 454 IEMANQPIVQITLVNEGCGYTGDIYFAKGDV-----LYSIELSGRIIL 500
Oy 466 EDHAQLTVNRFTQTCGVVSLGNGAVILSCYKNGAGNSASITLKHIGLSSILK--SGA 524
Db 501 REQTKLVNSLTQGTGSVHMEGSTLDF--AVTTPRANSMALTIVHESLASLTKNNGV 557
Oy 525 EIPILNVEPTN-----NSNNYTAOTAA--TFSLSDVKLSLIDYGNPSTLDTLHASSQ 578
Db 558 TNP-----PTNPVYVSSAVYIGNTAAGVTITSGP--IFFEDIDETAYANNOMLGADOTI 610
Oy 579 PMLISEASDNOLRSDMDPFGSLNVPHYGMGLWTMGACTODEPASPATITDPOKANR 638
Db 611 DVIQLHGANPPRANATDTLTLGNHSSKTYGQSWITQW----EPDA-----NPPQNN 660
Oy 639 FHRITLLTLMIPAGYVSPKRRSPLIANTLMGNMLLATESLKNSAELTPSDHPFW--GITGG 697
Db 661 Y--MLKASMTKTGYNGBPRVAVSLVNSLMS--ILDVRSASHAIQASIDGRVRCGIWIS 717
Oy 698 GICMAYVODPRENHPGFHMRSSYSAGMIA--GQTFSLKFSQTYTKLBERAKNNV-- 753
Db 718 G1SNFHYHODALGQGYRHISGYSIGANSYFSS--MFGIATFETIGR-----SKOYVVC 771
Oy 754 SSKNYSCEGEMFLSLOEGFLTKLVG-----LYSYGDHNCHEFTOGENLTSQGTFRSQ 807
Db 772 RSNDDHTCVGSVYLTQALOGSCLFCDATVRASYGCGNHMKSTYFAEE--SNVMDWN 829
Oy 808 TMGCAVFPDLPKPPGSTHILIT--APFLCALGIYSSLSHTEVAVPRSPSTKPLINLV 865
Db 830 CVYGEVAGAGLPIMLASKLYLNLRFVQAEFVYAEHSESTERGDOARFEKS--GHLMN 888
Oy 866 VPIGVKGSFMNATFORQAMTVELAYQVLYROBPGLATOLLASKGIMFGSGSSRHANS 925
Db 889 IPGVGFVD--RCSSKHPNKSIFEMCAIYCDARISIGETITLLSHKLTWTDAFLHAGVA 947
Oy 926 YKISQOTPLSMTLHFQYHG 946
Db 948 VRGSMYAS---LTGNIEVYG 964

RESULT 14
O9RB66 PRELIMINARY: PRT: 930 AA.
ID 09RB66:
AC 09RB66: 01-MAY-2000 (TREMblrel. 13, Created)

```

```

Db 713 LMGS-IIDIRSAHSAIQASVDCRSYCRGLAVSGVSNFEYHCRDALCGGYRYISGYSIG- 770
Oy 726 IAGOTHTFSLKFSOTYTKLNERAKNNV--SSKNVSCGEMLFSLQEOGFLTKLVGLSY 783
Db 771 --ANSYFSSSMFGLAFTEVFGFR-SKDYVCRSNHHCAGISGYLSTKQ-----LCGSTLF 822
Oy 784 GDHNCNHHFYTOG-ENLTSGTFRSOT-----MGAVFEDLPMPKPGSTHILT--APF 832
Db 823 GDAIFRASYGNGCHMKSTYFAESDVRMDNCLVGLIGLPIVITPSKYLINELRPF 882
Oy 833 LGALGIYSLSHFTEVGAVPPSFSTKTPPLINVLVPIGVKGSFMNATORPOAWTVELAYOP 892
Db 883 VQAFPSYADHSEFTFEGDQARAFRS--GALMLNLSYVGVKFDRCSSST--HPNKYSFMGAYIC 940
Oy 893 VLYRDEPGIATOLLASKGIWFGSSSPSRHAMSYKISQOTOPLSMLTLHFQYHG 946
Db 941 DAYRTISGQTTLTLLSHOETWTTDAFHLRHGVIRGSMYAS---LTSNIEVYG 990

```

RESULT 10

```

O92898 PRELIMINARY: PRT: 936 AA.
AC 092898:
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DE 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY.
GN PMP-7.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN NCBI_TaxID=83358;
RP SEQUENCE FROM N.A.
RC STRAIN=CWLO29;
RA Kalman S., Mitchell W., Marathe R., Jammel C., Fan J., Hyman R.W.,
  Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
  "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
  Nat. Genet. 23:385-389(1999).
RL EMBL: AF001627; ADL18589.1;
SQ SEQUENCE 936 AA: 100079 MW: 8881078530194FC CXC64:

```

Query Match 12.7% Score 645; DB 2; Length 936;
 Best Local Similarity 27.0%; Pred. No. 9, 9e-36;
 Matches 262; Conservative 146; Mismatches 392; Indels 172; Gaps 44;

```

Oy 75 GAATTTDYLSFDTQKECIYFAKNLTPESGAGY-----ASPNSPVEIR 121
Db 54 GTTYSLSLSDVSFQNALGIPILASGCFLEAGDGLTFQGNHAKFAFINAGSAGTAVST 113
Oy 122 DTGPIVFEFN-----TCRPFSTSNPRAANRKIREGAIHAAOULYHNHNDVYGFEM 174
Db 114 SAADKMLFDFNDFRSLIISCPULLSPGOCALKSV-----GNLSLGNOSQIIFTO 164
Oy 175 NFGYVGGAISTANTFVSNOSCFLEMDNICIQTNTAGKAIYA-GTSNSESNNCDL 233
Db 165 NFSSDNKGVIINTKN-FLLSSTQSPASFSRN---QAFTRKKGCVYVYAACTTITENSPOIVS 220
Oy 234 FTINMACAGCAIFS-PICSLTGNRCNIVFYNNRCRKNVETASSEASDGAIKVTT--R 289
Db 221 FSNGLMAGSGALYSTNCISITDN-EGVIFDGSAMWEAAQ-----AQGALICCTTIDKT 273
Oy 290 LDTVGNRGRIFESDNTKKNKGAIYAPVTVLVNPGPTFFINNIAANK-----GGAIIIDG 344
Db 274 VTLTGNK-NLSFTNNTALTYGALISGLKVSISAGGPTLFFQSNISGSSAGCGGGAIIINAS 332
Oy 345 TSNKISANDRAHIAFENENITVNTNANGSTSANPPRRNATVAVSSSGEILLGAGSSQNL 404
Db 333 AGELALATSQDITFNNQVNTN-----GSTF-----RNAINIIDTKVYSIRAAATGOSI 382
Oy 405 IFVDPLEIVSNAGVS-----FNKEADQOTGSVVFSGATVNSAD-FHQRNLQTKTP 453

```

```

Db 383 YFYDPI--TNFGTAASDTLNLNLADANSELEYGAIVFSEKISPEPKALAAVNTSTIR 440
Oy 454 APLTSLNGFLCIEDHQAOLTVNRFTQGVVSLGNGAVLSCYKNGAGNSASNASITLKHIG 513
Db 441 OPVALRARDLVLRGVVYTFEPKDLTGSPGSRILMDC-----CTTILSAKLANLSLNGIA 492
Oy 514 LNLSSILKSGAEIPLLVNEPPIVNSNNTADTAATFSLSDVLSLIDYDGNSPYSTDLTH 573
Db 493 VNLSL-----DGTNKALKTEADKNISLSGT-IALIDEG-SFENHNILKS 538
Oy 574 ALSSOPMLISEASON-OLRSDMDPFGSLNP--HYGOGWLTGMWATOPPEPASSATI 630
Db 539 A-STYPLLELTAGANGTITGALSTLTLOEPETHYGOGMOLSMAN-----ATSKSI 591
Oy 631 TDPOKANRFRHTLLTWLPAGVYVPSPKIIRSPLIANTLWGNMLLATESLKSNAELTPSDHP 690
Db 592 GS-----INWTRGTGIPSPERKSINPLNSLWGN-FIDIRISINQLETIKSSGEP 638
Oy 691 F-----WGITGGGLGMVYODPRENHPGFHMSSGYSAGMIAGOTTFSLKFSQYTXLNE 746
Db 639 FERELM-----LSGLANFERYDSMPTRHGRHISGVALGITAATPAEDQLTA--FCQLFA 693
Oy 747 RYAKNNVSSKN-----YSCGEMLFSLQEGFL-----LTKLVGL----- 780
Db 694 R-DNRNHTIGKNHGDYTGASLFFHTEGLFDI-ANFLMKATRAPVLSLEISQIITPLSPDA 751
Oy 781 --YSYGDHNCNHHFYTOGENTLSQGTFRSQTGAVFEDLPMPKPGSTHILT--TAPFLGA 835
Db 752 KFSYLAHTDNHMKTYT--DNSILKGSWRNDARFADLCAISLTA-VISVYVLLKEVPEPVKV 808
Oy 836 LGIYSSLSHFTEVGAVPPSFSTKTPPLINVLVPIGVKGSFMNATORPOAWTVELAYOPVL 894
Db 809 QYIVAHQDFEYRIHAEGRAF-N-KSELINVEIPIDVY---FERDSSSEKGTIDLTMLYLD 864
Oy 895 -YRDEPGIATOLLASKGIWFGSSSPSRHAMSYKISQOTOPLSMLTLHFQYHGYGFSSTF 953
Db 865 AYRRNPKCOTSLIASDANMAYGTNLAROGFSYRAANHQVPPHHEIGQF-AFEVRRSS 923
Oy 954 CNYLNGELALRF 965
Db 924 RNV-NINLGSKF 934

```

RESULT 11

```

O93542 PRELIMINARY: PRT: 936 AA.
AC 093542:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DE 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (POLYMORPHIC MEMBRANE
  PROTEIN G FAMILY).
GN PMP-7 OR CP0308.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN NCBI_TaxID=83358;
RP SEQUENCE FROM N.A.
RC STRAIN-J138;
RA Shitai M., Hatake H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
  Shiba T., Ishii K., Hattori M., Kihara S., Nakazawa T.;
  "Comparison of whole genome sequences of Chlamydia pneumoniae J138
  from Japan and CWLO29 from USA.";
  Nucleic Acids Res. 28:2311-2314(2000).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AR39;
RA MEDLINE=20150255; PubMed=10684935;
  Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
  White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,
  Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,

```

OY 960 EIALRF 965
Db 941 GTTFKF 946

RESULT 7

O92880 PRELIMINARY: PRT: 946 AA.
AC 092880.
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN E/F FAMILY.
GN PMP-18
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83358;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CML029.
RX MEDLINE=99206606; PubMed=10192388;
RA Kallan S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RT Nat. Genet. 21:385-389(1999).
DR EMBL: AE001631; AAD18610.1;
SQ SEQUENCE 946 AA: 103655 MW: C98B904E863C452C CRC64:

Query Match 18.6%; Score 948.5; DB 2; Length 946;
Best Local Similarity 29.1%; Pred. No. 1.9e-56;
Matches 281; Conservative 167; Mismatches 417; Indels 101; Gaps 34;

OY 35 PKESLSNKLSTGDTNLT-----NCYLDMLRYLALQTPNEGAVTTIDVL 84
Db 47 PLIDLTN--MTPYSHRATLFCGRDODTODIYLDHONSIESWFENFSOGGALSCS-L 102
OY 85 SFEDTQKEGIFYFAKNLTPESGAIGYASPNSTVEIRDTGPIYFENNTCCRFSTSNPN 144
Db 103 AITNT-KNQLFLNSFAIKRAGM-YVNGN---FDLSENIIGLTFGSLN-----SFPN 150
OY 145 AA--VVKIREGGAIAHONLTYNNHNDVYCFMKNFSYRGCAISTANTFYVSENQCFLEM 202
Db 151 ASNFADTCTGVALCSNVTISKNGAYFINKKAKSSGGAIOAIIINKDTPCLFF- 209
OY 203 DNICIGTNTAG--KGAITAGTSNFSNCCDLFINNACCAGAT-FSPICSLTGRGN 259
Db 210 -----NNAAGGTAGGALFANACR-IENNQPIYFLNOSGLGALRVHQECILTKVTGS 262
OY 260 IVEYNNCFNVEFASSEADGAIKVTTRLDYTGNGRIFFSFNITKKNGAIVAPVT 319
Db 263 VIFNN--FMEDDISANHSSGAIYIS-CSIKDNGIAFAFNNTAROGAICQTSLT 319
OY 320 LVNGPTFYFNINANNKGAIYIDGTSNKSISADRAHAIIFENIIVTVNANGTSTSANP 379
Db 320 IQDSGPYFTNN--QGTWGAIMLRQDCACTLFADQGDILFYNNHFRDFTSN--HVSVC 376
OY 380 PRNNAIVASSSGEILLGAGSNNLITDPL----EVSNAVSSTFKKEDQGSVYFSG 435
Db 377 TRNYSLVGASG-----HSATFYDPILOYRTIONS--IQKFPNDEHLOTILFSS 425
OY 436 -----ATVNSADF--HORNLOTKPAPULTSNGFLCIEDHQLTVNFTQTGVVSLGNG 488
Db 426 TYIDTSTSRDFTSHRN-----HGLYNGTLALEDRAEMKYYKFDGCGGLRLIGSR 478
OY 489 AVLSCKYKAGNSASNAITLKHIGLNLSSILKSGAEIPILLVEPTNNSNNYATDAATF 548
Db 479 AVFSTTDEGSSSVGVININNLAINLPIL-GNRVAPKIMIRPTGSSAPYSEDNPII 537
OY 549 SLSDVKSLIDYDGNSPRESTDLTHALSSOPMLSTSEASDQSLSDMDDFGSLN-VPIYG 607
Db 538 NLSG-PLSLDDENLDPYDADLAQPIAEVPLLYLLDTAHHINTDNEYPEGTLTQHYG 596

OY 608 WQGLMTWGMKATODPEPASSATITDPOKANRFHRTLLTWLPAGYVSPKRRPLANTL 667
Db 597 YGVWSPYIEITITTSOTSS-----EDVTNLTLRQLYGDWPTGKYVPEKKGGIATSAF 651
OY 668 W---GNMLLATESLKNSAELTPSDHPFMCITGGGIGMNVYODPENHNGFMRSSGSAG 724
Db 652 WQSPHNLFLATLRYOTOGCOIAPT-----ASGEATRLVHONSNDKAFMEATGSLG 705
OY 725 MIA--GQHTFSLKFSQTYTRKLNERYAKNNVSSKNVSCGEMLES-LOEGFLTLKLYGLY 781
Db 706 TTSNTASNSHSGVNSOLFNSNYESHSDNSVASHTTTVALQINNWDLERSTASLA-Y 764
OY 782 SYGDHCHHFTYTGNN--LTSOGTFRSOTMCGAVFEDLPKMPFGSTHILTLAPLCAIGY 839
Db 765 SYSN--HHIKASGYSKRIQTEGKCYSTLLGALSCSLQ-WRSRPLHFTPFIOALIVR 820
OY 840 SLSHFTVGANPRSFSTKTPILNVLPVIGVCSFMNATOPROATVELAVOPVLYROEP 899
Db 821 SNOTAFQESGDKARKFSYHKPLYNLTPLUGIOSAWESKFLPTYNILATOPVLYXQNP 880
OY 900 GIATOLLASKGIMFGSSPSSRHAMSYKISQOTPLSMILLHFYHGFYSSSTPCNYLNG 959
Db 881 EIVNSIESGSSWLLSGTLANALATAFGRNOLFIPKLSVLYDVGSSSTTHYLHA 940
OY 960 EIALRF 965
Db 941 GTTFKF 946

RESULT 8

O9PL46 PRELIMINARY: PRT: 1025 AA.
AC 09PL46;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE POLYMORPHIC MEMBRANE PROTEIN E/F FAMILY.
GN TC0262.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / NIGG;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C.,
RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,
RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia pneumoniae AR39.";
RT Nucleic Acids Res. 28:1397-1406(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Read T.D., Brunham R., Shen C., Gill S.R., Heidelberg J.F., White O.,
RA Hickey E.K., Peterson J., Umayam L.A., Uterback T., Berry K.,
RA Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,
RA Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G.,
RA Salzberg S.L., Eisen J., Fraser C.M.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE002293; AAF39131.1;
DR TIGR: TC0262;
SQ SEQUENCE 1025 AA: 11458 MW: EA472E07818B9FE2 CRC64:

Query Match 15.6%; Score 793; DB 2; Length 1025;
Best Local Similarity 27.0%; Pred. No. 8.9e-46;
Matches 269; Conservative 158; Mismatches 419; Indels 152; Gaps 35;

OY 68 LQKTPNEGAVTTIDYSPFDQKEGIFYFAKNLTPESGAIGYASPNSTVEIRDTGPIV 127

ID	992882	PRELIMINARY:	PRG:	934 AA.
AC	092882			
DT	01-MAY-1999	(TREMBLrel. 10, Created)		
DT	01-MAY-1999	(TREMBLrel. 10, Last sequence update)		
DT	01-OCT-2000	(TREMBLrel. 15, Last annotation update)		
DE	POLYMORPHIC OUTER MEMBRANE PROTEIN E FAMILY PMF_16 PROTEIN.			
GN	PMF_16.			
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).			
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.			
OX	NCBI_TaxID=83558;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-CML029;			
RX	MEDLINE=99206606; PubMed=10192388;			
RA	Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,			
RA	Oliver L., Gimwood J., Davis R.W., Stephens R.S.;			
RT	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis."			
RL	Nat. Genet. 21:385-389(1999).			
DR	EMBL: AE001631; AADI8609.1;			
SO	SEQUENCE 934 AA; 102149 MW; E88C8F6B11D13C47 CRC64;			
Query Match	20.7%;	Score 1052;	DB 2;	Length 934;
Best Local Similarity	31.2%;	Pred. No. 1,6e+63;		
Matches 296;	Conservative 158;	Mismatches 431;	Indels 64;	Gaps 27
35	PKTESLNK-----ISLTGDTFHNLT-NCYLDLRYLAILLOKTPNEGAAVITTDYLSFF	87		
32	PLADSLSTLTGSPITPTTYGNGRNSQDILVSKSIDNILLMTSAGCAVSCNFLT-L	89		
88	DNOKEIYFAKMLTPESGALGYASPNSPTVEIRDTTGYVIEFNNTCCRFPTSSNNAAV	147		
90	SNVEDHAFESKNIATIGTGALACQG---ACITTKNGRLPIFFSNRGLN-----NAST	138		
148	NKREGGAIHAO-NLYINHHDDVVGPMKNFYSYRGGAISRTANFVVSSENSOCLEFMDNIC	206		
139	GGETRGCAIACNCDFTISNOGTFYFVNNSVNMWGGALSTNGCHRIOSNAPLFFN---	195		
207	IQNTAGKGAIGAIAGTSNSPFESNNCDLFFINNACAGALFSPI-CSLTGNGNIYFYNN	265		
196	---NTAPSGGALRSEHTTISDNTRPIYFNKNGCNGNCAIOTSVAIVAKINNGSVIFENN	252		
266	RCKNVEITASSEASDGAIVKTRLDVTCGRGRIFPSDITTKNGCAIYAPVYTLVNDGP	325		
253	TALSG--SINGSGSGAI-YTNLNSIDDPGTLFFENNVCIRGGALICQTFLLTKRSH	309		
326	TYFTNNIANNKGAIVYIDGTSNKSISDRHAIIFENNIYVNVINANGSTISANDPRNAI	385		
310	VFTFTN-OGWVGALMLLDSTCLLRAEGQNIAPONNEVFLTTG-----RYNAI	358		
386	TVASSGEILLGAGSSONLIFYDPIEVSNGVS-VSEPKKADOTGSYVFSGATVNSADFH	444		
359	HCPNNS-NLDGANKKGYTAAFPDIEHQHTPTNLIENPNANHGTILFESSAYIPREASDY	417		
445	QRNLQTTTPAPLTLISNFCLEIDHAOLTVNRFTOTGGVSVLGNAGVAISCKYKNGAGNSA-	503		
418	ENNNISSKSTISELRNGVLSIEDAGWQFYKFTQKGGLIKGHAASIAITTANSETPTSV	477		
504	NASTFLKHIGLISLILKSGAEIPLWVEPTNNNNYTAADAAAFSLSDVLYSLIDDYGN	563		
478	GSQVITNNLAINLPSILAKG-KATTLWIRLPLOSSAPFTEDNNPITITLSG-PLTLNENR	535		
564	SPYESTDLTHALSSQPMLSISEASDNDLRSDMDPSGLN-VPHYGWGLTWGAKATDP	622		
536	DVYSIDILSEPLQNIHLTISLSDVTAHINTDNFPESEINATEHYGYOQIWSPYWET--I	593		
623	EPASSATTDPOKANRFRITLLMLPAGVYDPSKRNRPPLANTLWGMMLATESLKRSA	682		
594	TTTNNAST---ETANTLRYANKPTPLGYVNVDEYQDCLTTPLMQSFHMFSLSRYN	650		
683	ELTPSD--HPWGTITGGCIGMNVYODPRENNRPFGRMSGYS--AGMLAGOTHTFSLKFS	738		

Query Match	20.6%	Score 1046	DB 2	length 934
Best Local Similarity	31.1%	Pred. No. 4.2e-63		
Matches 295	Conservative 158	Mismatches 432	Indels 64	Gaps 27
35	PTKESLSNK-----ISLGDTHNLT-NCYIDNLRKYLIALQKTPDEGAAYITDVLSPF	87		
32	PLDSDSLTLTGFSPIITTFVGNRHNSODIYLSYKSIDNLTMLMTSAGVAGSCNNPL-L	89		
88	DIOKEGTFEKNLTPESGGAIGYASPSFVEYRDTIGPVIYFENNCCRPETSSNAAV	147		
90	SNVEDHAFSKNLAIAGTACAIACOG-----ACTLTNRGRLPTEFSNGLN-----NAST	138		
148	NKIREGAIHAQ-NLYINHNHDVYGEKKNESYVYGAISTANTFVYSENSCFLPDNDIC	206		
139	CGEIRGAILNCIDPFTISQNOGTFEYFVNNSVNMVGGALSTNGHCRIQSNRAPLFFN---	195		
207	IQNTACKGCAIYAGTSNSFESNNCLDFTFINNACGACAFSP1-CS1JCNRCNIYFNF	265		
196	---NTASGGGALRSEHTTISDNTRP1YFKNNGCNGNAGAIQTSYVAIKKNSGVIFENN	252		
266	RCFNNVETASSEASDGAIVYTRFLDVTGKRGRIEFSNDITKKNYGAIVPYVTLVNDGP	325		
253	TALSG--SINGSGSGAI-YTTNLSIDDPRTLLFNNNVCIRDGALICITOPLTINSGH	309		
326	TYFINNIANNKGAIIYIDGTSSNKSISADRIHA1IFENIYTVTNANGTSTSNPRRNA1	385		
310	VYFINN-OGMMGKMLMLDOSTCLLPEDGQIARONNEVEFLTTG-----RYNA1	358		
386	TVASSGCEILGACSSONLIFYDIEVSNAGV-SVRNKEADDTGSGVPSGATVNSADFH	444		

```

Db      1 MKKLEFFVLIGSSILGTFREVPPSLLKRLPLNPFYHMTGLFFPKVNLIGDTHNLTGYHLDN 60
Oy      61 LRYTLAILOKTPNEGAATITDYLSFPDQKGIYFAKNLTPESGAGYASPNPTVEI 120
Db      61 LKCLLACLRPVECAALFVTDYLGFSDDQKGIYFAKNLTPESGAGYASPNPTVEI 120
Oy      121 RTIGPVIFFENNTCCRPFTSSNPNAAVNRKREGGAIHAQNLINHHHDVYGFKNKESYVR 180
Db      121 HNTIGPVIFFENNTCCRPFTSSNPNAAVNRKREGGAIHAQNLINHHHDVYGFKNKESYVR 180
Oy      181 GCAISTANTFVVENOSCFLEMDNICIOTNTAGKCAIYAGTSNFSFNCDLFLINAC 240
Db      181 GCAIS-ANTFAKEMKSSFLCLNNSCICQTKGCAIYVSTSGENNKKDLFLIONSG 239
Oy      241 CAGGAIFFSPICSLTNGRGIYFYNNRCFKNVETASSEADGAIKVTTRLDYTGNGRIF 300
Db      240 CAGGAIFFSPICSLTNGRGIYFYNNRCFKNVETASSEADGAIKVTTRLDYTGNGRIF 299
Oy      301 FSDNTKNGCAIYAPVTVLVNCPPTFYFINNANKGCAIYIDGTSNKSISADRAHILFN 360
Db      300 FSDNISRNGGAIHAPCLHLVNGPFTYFNNTANITGCAIYITGTETSKISADHILITFD 359
Oy      361 ENIVTNVNTANGSTSANPPRRNAITVASSGEILLGAGSSONLIFYDPIEVNAGVSVS 420
Db      360 NNISANATNADGSSNTNPPHRAITMDNSAGCIELGACKSONLIFYDPIQVTVNAGVTVD 419
Oy      421 FKKEDQGSVYFSGATVNASDFHORNLOTKTPAPLITLSNGFLCIEDHQLVYNRTQVG 480
Db      420 FKKDASQTCVYFSGATVNASDFHORNLOTKTPAPLITLSNGFLCIEDHQLVYNRTQVG 479
Oy      481 GVVNSGNGVILCYKNGACN-----SASNAITLKHIGNLSSILSKCAEIPLLWVE 532
Db      480 GIVAAAGNVISYOHSTIDATOTPTTTTDDASVTLNIGLNLPSILDGAEMPLIWE 539
Oy      533 PTN---NSNNYTAADTAATFSLSDVNLSLIDDYNSPYESTDLTHAALSQPMLSISEAD 588
Db      540 PLSTTGNTTTSYDPAASFSLNGATLSLIDEDGNSPYENTDLSRALYQPMALISEAD 599
Oy      589 NQLRSDMDPESGLNPHYKNGOGLTGMKAKTODPREPASATITDPOKANRPHRTLLTTL 648
Db      600 NQLOESMDPESGLNPHYKNGOGLTGMKAKTODPREPASATITDPOKANRPHRTLLTTL 659
Oy      649 PAGYVESPKHRSPLANTLTMGNMLLATESLKNSAELTSPDHPHMGITGGGLGMVYQDR 708
Db      660 PAGYIYSPKHSPLANTLTMGNMLLATESLKNSAELTSPDHPHMGITGGGLGMVYQDR 719
Oy      709 ENHPGFHMRSSGYSAGMIAGOTHTFSLKFSQTYTKLNERAKNNVSSKNYSCQCEMLFSL 768
Db      720 KHPGPHMHTSGYSAGMIAGOTHTFSLKFSQTYTKLNERAKNNVSSKNYSCQCEMLFSL 779
Oy      769 QCEGFLITKLVGLSYGDHCHHFTYTOGENULTSOGTFRSQTMGAVFEDLPKMPFGSTHLL 828
Db      780 QCEGFLITKLVGLSYGDHCHHFTYTOGENULTSOGTFRSQTMGAVFEDLPKMPFGSTHLL 839
Oy      829 TAPFLGALCISLSHFTFVGAAPRSFTKTPILINLVYIGVSGPMNATOPROATVEI 888
Db      840 TAPFLGALCISLSHFTFVGAAPRSFTKTPILINLVYIGVSGPMNATOPROATVEI 899
Oy      889 AYOPVLVYRQEPGIAIOLASKGIWEGSGSSPSRHANSYKISQOTPLSLWLTJLHFOYHGY 948
Db      900 AYOPVLVYRQEPGIAIOLASKGIWEGSGSSPSRHANSYKISQOTPLSLWLTJLHFOYHGY 959
Oy      949 SSSITPCNYLNGEALRFE 965
Db      960 SSSITPCNYLNGEALRFE 976

```

```

RESULT 2
O92883 PRELIMINARY: PRT: 938 AA.
AC O92883:
DT 01-MAY-1999 (TREMblrel. 10, Created)

Query Match      22.0%; Score 1120.5; DB 2; Length 938;
Best Local Similarity 30.9%; Pred. No. 3.4e-68;
Matches 308; Conservative 163; Mismatches 418; Indels 109; Gaps 31.

Oy      6 FFLIGNSLSGLAREVSPHFLMPSVDPPTKESLSNKLSTLGTGDTHNL-----NC 56
Db      12 FTFVLAN--EGQLPLETYITLSPEVOAPQ-----GFTHNQMDLAIYGNHND 59
Oy      57 YLDNLRYLAILOKTPNEGAATITDYLSFPDQKGIYFAKNLTPESGAGYASPNPTVEI 116
Db      60 FLIDRYVY-----RSNGALTCNKL--ISENIGNVFEXNVCPSGAL-YAQNOC- 108
Oy      117 TVEIRDTIGPVIFFENNTCCRPFTSSNPNAAVNRKREGGAIHAQNLINHHHDVYGFKNK 176
Db      109 -----TIK--NQNATFTNLVSDNPTTAGSL--GGLFLFINCSITNNLGGQFVVDL 159
Oy      177 STVRGCAISTANTFVVENOSCFLEMDNICIOTNTAGKCAIYAGTSNFSFNCDLFLINAC 236
Db      160 ALNKGCAIYTFNLISIKDKGPIIHKONRAL--NSDSLGGIYSGNSLNEGNSGAIQIT 217
Oy      237 NNAACGGAIFG--PICSLTNGRGIYFYNNRCFKNVETASSEADGAIKVTTRLDYTGNGR 295
Db      218 SNSSGGGGIFSTQTLTJSSNKKLIEJSENSAFAN--NYGSNPNRGGGLTTTPTCTILNN 275
Oy      296 KGRIFSDNITKNYCAIYAPVTVLVNCPPTFYFINNANKGCAIYIDGTSNKSISADRAHILFN 352
Db      276 REGVLEFNNGOSNGCAIHAKSIIKENGPYFLNNTATRGALLLWLSGNSGNSGIFSLA 335
Oy      353 DRHAIIFENIVTNVNTANGSTSANPPRRNAITVASSGEILLGAGSSONLIFYDPIEVN 412

```


N100_YEAST
ID N100_YEAST STANDARD: PRT: 959 AA.
AC Q02629;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE NUCLEOPORIN NUP100/NSP100 (NUCLEAR PORE PROTEIN NUP100/NSP100).
GN NUP100 OR NSP100 OR YKL068W OR YKL336.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=93054906; PubMed=1385442;
RA Wente S.R., Rout M.P., Blobel G.;
RT "A new family of yeast nuclear pore complex proteins.";
RL J. Cell Biol. 119:705-723(1992).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=94378724; PubMed=8091863;
RA Rasmussen S.W.;
RT "Sequence of a 20.7 kb region of yeast chromosome XI includes the
RT NUP100 gene, an open reading frame (ORF) possibly representing a
RT nucleoside diphosphate kinase gene, tRNAs for His, Val and Trp in
RT addition to seven ORFs with weak or no significant similarity to
RT known proteins.";
RL Yeast 10:569-574(1994).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF
CC THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.
CC -1- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX.
CC -1- DOMAIN: CONTAINS G-L-F-G REPEATS.
CC -1- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z15035; CAI78753.1;
CC EMBL: X75780; CA53406.1;
CC FMBL: Z28068; CA81905.1;
CC PIR: B44402; B44402.
CC PIR: S39173; S39173.
CC PIR: S44518; S44518.
CC SCD: S0001551; NUP100.
CC NUCLEAR protein; Transport; Repeat.
CC FT DOMAIN 33 571 29 X 6 AA APPROXIMATE REPEATS OF
CC G-L-F-G.
CC FT SEQUENCE 959 AA; 99988 MW; D3985F9901BBA51 CRC64;
Query Match 2.8%; Score 141.5; DB 1; Length 959;
Best Local Similarity 19.8%; Pred. No. 0.3;
Matches 151; Conservative 81; Mismatches 257; Indels 273; Gaps 38;

QY 243 -----GGAI-F--SPI--CSLTGNR-----GNIVFYNNRCFKNVETASSE 277
DB 223 GQGSQNTNNAFGNQNLGSSFGSKFVGSGSLFGQSNNTLGTNTNRNGLFGQMSNQC 282
QY 278 ASDGAI-----KVTRLDYVGNRGRIFF--SDNTKNGCAIYAPVTVLV 321
DB 283 SSNSGLFGQNSMNSSTQVFGQNNQMOIINGNNNSLFGKANTFSASAGLFGQ--NNQ 340
QY 322 DNGPTVFINNIANKGCAIYIDCTNSK-----I 350
DB 341 QQSGSLFGQ--SQTSGSSQLFGQNNQKQPNFTQSNITGLFGQNNQOQSGTGLFGAK 398
QY 351 SADRHAIIENENIVTNTVANGSTANPPRRVAIVASSGEILLGAGSSONLIF--YD 408
DB 399 PACTTSSLFGQNSSTQPNLSFGTT--NWPFSN--TQSGQNSLFGATKLTNMPGCGNP 452
QY 409 PIEVSNAGVSVFENKADDTGVSFSGATVNSADFHQRNLQTKTPAPLTLNGLICIEDH 468
DB 453 TANQSGNSLFGTKPASTGSL-----FGNNTASTVPS--TNGL--FGNN 495
QY 469 AOLTVNRFTDTG-----GVYSLGNGAVLSCTKNGAGNSASNSI-----TL 509
DB 496 ANNSTST-TNTGLFGAKPDSQSKPALGGGLF-----GNSNSNSTTIGQNKPVFGTT 546
QY 510 KHIGL-----NLSTILKSG-----AEPLWVEPTNN--SNVYTDATATFSLSDV 553
DB 547 QNTGLGATNTSSAVGSGTKLFGQNNNTLVNTGVNPPVNNITQNNLGLTAVPSLQQA 606
QY 554 -----KLSTLIDYGNSTPEST-----DLTHA--LSQPMLSISEASDQ 590
DB 607 PTNQLFSLKISLSPNSTN-PVKATTSKVYADKRNLSLSAVRLAKPLFASSNGDAK 665
QY 591 LRSDMDFSLNPHYGMGGLMTWGAKT--QDPEASATITTDQKANRFRITLLTW 647
DB 666 FOK-----MCKTLERSDRGSSNSTITDESSYLANSNDLL--- 700
QY 648 LPAQVPSPKHRSPILANTLWGNMLLATESKNSAELTFPSDH 689
DB 701 -----FDPDRKRLKHLV-----IKNNKLNVLNH 724
RESULT 25
YV96_YEAST STANDARD: PRT: 1117 AA.
ID YV96_YEAST
AC P53753;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHEICAL 121.1 KDA PROTEIN IN BIO3-HXT17, INTERGENIC REGION
DE PRECURSOR.
GN YNR067C OR N3547.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RX Duesterheft A., Floeth M., Fritz C., Heuss-Neitzel D.,
RA Hilbert H., Moestl D.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO YEAST YLR144C AND S.POMBE SPAC2303.10C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z71682; CA96349.1;
CC SCD: S0005350; YNR067C.
CC Hypothetical protein; Glycoprotein; signal.

```

RL Can. J. Microbiol. 38:193-202(1992).
RN [2]
RP SEQUENCE OF 1-313 FROM N.A., AND SEQUENCE OF 1-20.
RC STRAIN-ATCC 19089 / CB15;
RX MEDLIN-89008089; PubMed:3049545;
RA Fisher J.A., Smit J.K., Agabian N.;
RT "Transcriptional analysis of the major surface array gene of
RT Caulobacter crescentus."
RL J. Bacteriol. 170:4706-4713(1988).
RN [3]
RP CHARACTERIZATION.
RC STRAIN-CB15;
RX MEDLINE-98292737; PubMed-9620954;
RA Awram P., Smit J.K.;
RT "The Caulobacter crescentus paracrystalline S-layer protein is
RT secreted by an ABC transporter (type 1) secretion apparatus."
RL J. Bacteriol. 180:3062-3069(1998).
CC -1- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
CC OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -1- MISCELLANEOUS: THE CALCIUM BINDING OF THIS PROTEIN MAY MEDIATE
CC SURFACE ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A
CC SPECIFIC MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF062345; AAC38665.1; -
DR HSSP: P01549; ZMCM.
DR InterPro: IPR001343; -
DR Pfam: PF00353; hemolysincabind. 1.
DR PRINTS: PRO00313; CABDNCRPT.
DR Cell wall: S-layer; Calcium-binding.
DR INIT_MET 0
FT SEQUENCE 1025 AA; 98209 MW; AEC8B519820B1A5F CRC64;
SQ

```

Query Match 2.9%; Score 146.5; DB 1; Length 1025;
 Best Local Similarity 19.6%; Pred. No. 0.16;
 Matches 107; Conservative 73; Mismatches 245; Indels 121; Gaps 20;

```

OY 35 PKRESLSNKLSTLGDTHNTCLDNL-----RYLAILQKTPREGAAVTTDYLSFEDTQ 90
DB 72 PSNAGIDELFDSTRTNDLNDAYISKFAQDENRINFSINLAIAGAGATN-----FAAA 125
OY 91 KEGYFAKNLTPESGGAIGYASPSNPTVEIRDTIG-----PVIFENNTCCRPPT 139
DB 126 YGVGSVAQGVAAATYDKIICNAAVATAGVDAVAALFLSRQANDLYLTAFAVANT---PFR 182
OY 140 SSNPNAAVKKIREGGAIRHONLYINNHVDVGMKMFYVRG-GAISTANTFVSENOSC 198
DB 183 A-----AADIDLAVKALIGLILNAATVSGIGYATATAAAMINDLSDG 225
OY 199 FLEMDNIC-----IOTNTAGKGAIIYAGTSNFSNNCDLFFINNACGAGAFSP 250
DB 226 ALSTDNAAGVNLFTAVPSSGVSSTSLTTGTDTLTGTANNDTFVAGEV--ACAALITVG 283
OY 251 CSLTGNRGNIIVFY-----NNRCRKNVETASSESDGAIKVTTRLDVYGNRGRI 299
DB 284 DTLSSGAGTDVLMWQAAVATLPTGVITISIEFTMN--VTSGAITLNTSSVGTGLTA-- 339
OY 300 FFSNDITTKNGAIYAPVT-----LVNPGPYFIININNNKGAIIYDGT----- 345
DB 340 -----LNTNTSGA--AQYVYAGAGQNTATTAAQAANNVAVDGRANVTASTVSTGTT 392
OY 346 --SNSKISADRAHIIIFENIVTN--VTNANGSTSANPPRNATITVASSGCEILLGASS 401

```

```

DB 393 VGANSAAGTYSVSVANSSTTTTGAIAVTGTAVTVAQIAGNAVNTLLQADVTVGNSS 452
OY 402 ONLIYDPELVSNAGVS--FNKEADOTGSVFGCAIVNSADFHORNLOKTPAPLTL- 458
DB 453 TTAATVTTOTAAATAGATVAGRVNGAVTTITDSAAASATTAGKI-----ATVTLG 500
OY 459 SNGFLICIDHQLTVNRTFGGVVSLGNGAVLSCKYKNGAGSASNSTLKLHGLNLS 518
DB 501 STGAATIDSSALTTVN-LSGTGTSLSIGRGAL-----TATPTANTLTLMVNG 546
OY 519 ILKSGA 524
DB 547 LTTTGA 552

```

Query Match 2.9%; Score 145.5; DB 1; Length 863;
 Best Local Similarity 19.5%; Pred. No. 0.15;
 Matches 159; Conservative 91; Mismatches 276; Indels 289; Gaps 36;

```

OY 117 TVEIRDTIGVIFENNTCCRPFTSSNPNAAVKKIREGGAIRHONLYINNHVDVGMKMF 176

```



```

OY 593 ---SDMDPFGSLNPHYMGGLMTN-----GNAKTOD----- 621
DB 1185 KTGCGDASFTLGNG--GFDVLTGYEYVLKSDGNSNMWLTNDVKNPDPPIPNKRPDPKD 1242
OY 622 -----PEPASSATITDPKANKFHRTLLLTWLPAGVNSPKHR-SPLIANTLMGNL 672
DB 1243 KPDPNPKPDPPTPT-----PYPERKITPSTAVAL--NM- 1277
OY 673 LATESLXNSAEL-----TPSDHPFWGIT-----GGGL-----GMVY 704
DB 1278 AATLPVLFDAELNSIRERLINIKKASPHNNNVWGATYNNVTTDAGAGEQTLTGMTVG 1337
OY 705 ODPRENHGFHNRSSGYS-CHIAQOTHTFLKESQYTKLNERIAKNNVSSKNISCGE 763
DB 1338 IDSRNDIP-----EGITTLGAFMGYSHS-HIGF-----DRGHHGVS--YSLGY 1380
OY 764 MFSLOEGFLTKLVGLSYGDHNGHFEYTOGENLTSQTFPSQTMGCAVFPDLPKPRG 823
DB 1381 ASMEHESGFYLDGVYKLNFRKSNVAGKSSSG--AANGSYHNSGLGHI--ETGKRFTD 1435
OY 824 STHILFADPLGALGIYSSLSHPT 846
DB 1436 GWNMLTP-----YASLTGFT 1450

RESULT 17
CSG_METFE STANDARD: PRT: 593 AA.
ID CSG_METFE
AC P27374;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CELL SURFACE GLYCOPROTEIN PRECURSOR (S-LAYER PROTEIN).
CN SIGA.
OS Methanothermus fervidus.
OC Archaea: Euryarchaeota; Methanobacteriales; Methanothermaceae;
OC Methanothermus.
OX NCBI_TaxID=2180;
RN 111
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-42.
RC STRAIN-DSM 2088 / V245;
RX MEDLINE=91293115; PubMed=1712296;
RA Broeckl G., Behr M., Fabry S., Hensel R., Kaudewitz H., Blendl E.,
  Koenig H.;
  "Analysis and nucleotide sequence of the genes encoding the surface-
  layer glycoproteins of the hyperthermophilic methanogens
  Methanothermus fervidus and Methanothermus sociabilis.";
  Eur. J. Biochem. 199;147-152(1991).
RN 121
RP CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE=94086483; PubMed=8262914;
  Katcher U., Schroder H., Haslinger E., Allmaier G., Schreiner R.,
  Wieland F., Haselbeck A., Konig H.;
  "Primary structure of the heterosaccharide of the surface glycoprotein
  of Methanothermus fervidus.";
  J. Biol. Chem. 268;26821-26826(1993).
RN 131
RP FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
  OF PROTEINS WHICH COAT THE SURFACE OF THE CELL.
  - SUBCELLULAR LOCATION: CELL WALL. THIS ARCHAEA IS COVERED BY A
  S-LAYER WITH HEXAGONAL SYMMETRY.
  - PRT: PROTEIN CONTAINS ABOUT 10 N-LINKED GLYCANS WHICH CONSIST OF
  METHYL-MAN, MAN AND GALNAc RESIDUES IN A MOLAR RATIO OF 2:3:1.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC EMBL: X58297; CAA41230.1;
DR PIR: S16225; S16225.

```

```

KW Glycoprotein; Cell wall; S-layer; Signal.
FT SIGNAL 1 22
FT CHAIN 23 593
FT CARBOHYD 29 29
FT CARBOHYD 58 58
FT CARBOHYD 66 66
FT CARBOHYD 74 74
FT CARBOHYD 114 114
FT CARBOHYD 122 122
FT CARBOHYD 145 145
FT CARBOHYD 148 148
FT CARBOHYD 158 158
FT CARBOHYD 176 176
FT CARBOHYD 208 208
FT CARBOHYD 231 231
FT CARBOHYD 326 326
FT CARBOHYD 336 336
FT CARBOHYD 340 340
FT CARBOHYD 431 431
FT CARBOHYD 471 471
FT CARBOHYD 500 500
FT CARBOHYD 516 516
SQ SEQUENCE 593 AA; 65481 MW; E06466967B7611C CRC64;

Query Match 3.0%; Score 153.5; DB 1; Length 593;
Best Local Similarity 22.9%; Pred. No. 0.029;
Matches 119; Conservative 46; Mismatches 189; Indels 165; Gaps 28;

OY 92 EGIYFAKNLTPESGGA--IGVASPNSPYEIRDTIGPVIFENNTCCRPSTSPNAAVVK 149
DB 61 EMLVKNLSLTKNSGSAIINAVSSEKSTITK-----ANNWIEGF----- 101
OY 150 IREGG--AIHAQNL--YINHNDVYGFPMKNFSYVYGAI-----STANTF--VVSSENS 197
DB 102 IIGCGNGIYMEVNGCTITNTTIOAFVSGMEYVGNGICLVNSNNITNNIIRNTW 161
OY 198 CFLFDNIC-----IQTYACKG--AIYAGTSNFSFESNN-----CDLF- 234
DB 162 NGI---NVCESKNIIKNTIMYSGIGIYVWGFNFKFEGNNIENNRIINATYGCITYLR 218
OY 235 -----FINNACCAGAIIFSPIC-----SLTGNRGNTV 261
DB 219 PSNNKTCRNVIYAVSSGGGSGGSGAICIDVSDYNIYKNDIGICDGLFTDGMIGNE---- 274
OY 262 FYNNRCFKNVEVETASDGAIKVTRIDVYGNRGRIFFSDNIRKNYGAIYAPVYTLV 321
DB 275 -ITNIFKNCQVAVSESTYGPA-----SRNKIY--GNFYINVEFTALSDPGELV 321
OY 322 DN-----GPTFYNNIANNKGAIIYDGTSSKISADHRAIIFN-----ENIYV 365
DB 322 DNIMNTTEGGNWSNVTGNNTG-----DGTGNIPYYDMKPLVVDLAIEDIAKPSGLEV 376
OY 366 NVTNANGTSTANPPRRAITVAASGEIL-----LGAGSSONLIFYPIEVSAGVSV 419
DB 377 RKKNLGRADIKIDPLTKIKISCNDVYETIFIDPLSGESO-IYRWKI-VPEGNHTI 434
OY 420 SFNKREADQTSVVFSGATVNSADFHQRNLQTKTPADLTLSNGLIEDHQAQLVNRFTQ 479
DB 435 KAEIPYSAGYLI--GTNIRADI--SNVEFSK-----IVQGF-----YQKFTT 475
OY 480 GGVVSLGNGAV-----LSCYK--NAGNSASVASTLK 510
DB 476 IYLTNLGKSTIITIKYIYITPVNCTKVSYRELITLK 514

RESULT 18
CSG_METSC STANDARD: PRT: 593 AA.
ID CSG_METSC
AC P27374;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)

```

Best Local Similarity 21.4%: Pred. No. 0.059:
Matches 127: Conservative 78: Mismatches 233: Indels 155: Gaps 25:

```

OY 209 TWTACGCAIYAGTSPNSPESNCDLFFINNACAGAIFF--SPICS-----LTGNGRG 258
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 32 NSGSGGGG---GSLVSNSSNCG---SNNISRKSTLMYGNCPHSLINDGKELLENSKP 84
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 259 NIV-----FYNNRCEKVNVTASSE-----ASDGAIXKVTTLRDVYG----- 294
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 85 KYVELVNTFNHKKPLSTISHVNEIIPPAIEKEKEIINTISGVTAKYMTALEILDSTIN 144
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 225 ---NNGR---IFFSDNITNYGCAIYAPVLTVDNCPTEFINNIANNKCAIYIDGTSNK 349
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 145 TPLNRSRSGSISGKRICNNLTSSSSSSTTATPTPTSSNNNNNNNNNNNNNNNNNN 204
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 350 ISADRHAIIFENIVTNTNANGTST-----SANNPRNAITVASSGCEIILGAGS 400
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 205 NNNNNNNNNNNNNNNNNNNNTTTTITSLISSPPSS--SSSSNDECFNNNN 263
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 401 SONLIFYPIEVSAGVSVFKEADOTGVSFSGATVNSADFHQRNLOTKTPAPLTLSN 460
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 264 NNN-----SNSSGSSRMITSKSIKPLI---VTSNTAATTTTTTTTNTSAPTPTPN 311
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 461 GELCIEDHQAOLTVNRFOTGTGVSILGNCVAVLSCYKNGAGNS-----ASNASITLKHIGL 514
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 312 RVQSSLD--DLFLNLPITPSNVPTVNGPKISAVPKKVSCKLIPSSNVS--SSNITL 368
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 515 NLSTLKSAGAEIPLLWEPNTNNSNYTADTAATFSLDVKLSLIDYGNSPYESTDLTHA 574
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 SLSS-----SSPSSSSSSTSTVPIVQLS-----SNTSPSTSLPTTPR 410
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 575 LSSQMLISE---ASDQLRSDMDPFGSLNPHYGOMGLTWGAKVODPEPASATIT 631
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 411 L-SQPTTSTVQLIPSQOQOQPPESNSSNTNT-----TTSSSSSSSSSLTIS 458
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 632 DPQKNNR-----FRTLTLTLWLPAGYVP--SPKHSPLIANTLGMNMLAT--ESLKN-- 680
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 459 SQPNNNSIRISAFRRSSQTFISSNGIPSPGQVSNKYNNI--GNLSSSEERKNNXY 517
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 681 -----SALPLPSDHPFMCITGCGIG 700
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 518 SMLNISKRTILDESDISSPISGSPNISIRASISSQLPFLSSIGGGGGGGG 570
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 15
ID YDEK_ECOLI STANDARD: PRT: 1325 AA.
AC P32051: P76140: P77168;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHEICAL 136.5 KDA LIPOPROTEIN IN HIPB-UXAB INTERGENIC REGION
DE PRECURSOR (ORF1).
GN YDEK OR ORF1.
OS Escherichia coli.
OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
OC Escherichia.
OX NCBI_TaxID=562;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617: PubMed=9278503:
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
RL [2]
RN SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=9725157: PubMed=9097039:
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,

RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio T., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivassundaram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.,
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map."
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 595-1325 FROM N.A.
RX MEDLINE=94100243: PubMed=8274505;
RA Catwright P.J., Tims M.W., Lithgow T., Hoef P.B., Hoogenraad N.J.;
RT "An Escherichia coli gene showing a potential ancestral relationship
RT to the genes for the mitochondrial import site proteins ISP42 and
RT MOM38."
RL Biochim. Biophys. Acta 1153:345-347(1993).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
CC ANCHOR (POTENTIAL).
CC -1- SIMILARITY: TO E. COLI YFAL.
CC -1- SIMILARITY: SOME, TO FUNGAL MITOCHONDRIAL IMPORT SITE PROTEINS
CC ISP42 AND MOM38.
CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 653.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib.ch).
CC -----
DR EMBL: AE000248: AAC74583.1: -
DR EMBL: D90793: BAA15190.1: ALT_INIT.
DR EMBL: D90794: BAA15197.1: ALT_INIT.
DR EMBL: X73295: CAAS1730.1: ALT_FRAME.
DR PIR: S34315: S34315.
DR Ecogene: EG11780: ydek.
DR Prosite: PS00013: PROKAR_LIPOPROTEIN_1.
KW Hypothetical protein; Membrane; Lipoprotein; Signal.
FT SIGNAL 1 18
FT CHAIN 1 18325
FT LIPID 19 19 N-ACYL DIGLYCERIDE (POTENTIAL).
FT CONFLICT 884 884 N->K (IN REF. 3).
FT CONFLICT 1317 1317 M->S (IN REF. 3).
FT SEQUENCE 1325 AA: 136514 MW: 26A3A066FA19A07D CRC64;
Query Match 3.1%: Score 157.5; DB 1: Length 1325;
Best Local Similarity 18.6%: Pred. No. 0.049:
Matches 177: Conservative 113: Mismatches 346: Indels 315: Gaps 40:

```

OY 108 IGVASPNSPYIEIRDTICPVLENNTCRPTSSNPNAVNRKREGCAIHAQNLTYNNH 167
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215 LGYQAGSNQVVV--EKGEWLIKNNDSIEFQIGNGTGEATIRGGLVLAVENTIIGNA 273
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 168 DVGGFMK-----NESYRGAISTANTFVVSENO--SCFLPMDNLCIOTNAG 213
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 274 TGIGTLNVOODSDVITRRLVLYNGYFGNGTGNISNGLLNKKRYSLVGVOGDSHGVVNTD 333
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 214 KCGAIYAGTSNFSFNENCDLFFINNACAGAIFFSPICSLTGNRCNIYFYNNRCKVET 273
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 334 KCHNMFLLGIGEAFF-----YIYIGDA-----GDEI-----N 360
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 274 ASSEAS--DGGAIKVTTRLDVTCNGRIFFSDNITNNGAIYAPVLTVDNCPYFINNI 332
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 VSSEKQVDSGIIITACMKETGTN-----ITYDK-----NSV 392
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 333 ANKCGAIYIDGTSNKSISADRHAIIFN-----ENIVTNTNANGTSTSNANPRRNA 384
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 393 ITNLGTNLGYDGHGEMNIS--NOGLVSNCGSLG/GETGVGVNSITTTGGMEEVNNVYTT 451
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 385 ITVASSGCEIILGAGSSONLIFYDPIEVSAGVSVFKEADOTGVSFSG-----ATV 438
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


DR	EMBL: Z48544; CA688442.1; -
DR	Mormped: ZK945-9; CE01740.
DR	InterPro: IPR000203; -
DR	InterPro: IPR001024; -
DR	Pfam: PF01825; GPS: 1.
DR	Pfam: PF01477; Pfam: 1.
KW	Hypothetical protein. Repeat: Transmembrane.
FT	DOMAIN 22 115
FT	250 389
FT	TRANSMEM 557 577
FT	TRANSMEM 606 626
FT	TRANSMEM 1162 1182
FT	TRANSMEM 1290 1310
FT	TRANSMEM 1467 1487
FT	TRANSMEM 1498 1518
FT	TRANSMEM 1541 1561
FT	TRANSMEM 1602 1622
FT	TRANSMEM 1637 1657
FT	TRANSMEM 1717 1737
FT	TRANSMEM 1935 1955
FT	TRANSMEM 1990 2010
FT	TRANSMEM 2039 2059
FT	TRANSMEM 2088 2108
FT	TRANSMEM 2134 2154
FT	TRANSMEM 2169 2209
SO	SEQUENCE 2339 AA; 254348 MW; 4D9DD23060FOCA85 CRC64

Query Match	3.4%;	Score 170.5;	DB 1;	Length 2329;
Best Local Similarity	21.1%;	Pred. No. 0.017;		
Matches 208;	Conservative 140;	Mismatches 428;	Indels 211;	Gaps 42

```

0Y 20 EWP5SIFLMP5VDPDPKRESLSNKKISLTGDPHNLNCTYDNLURYLIALLOKTPNEGAAYT 79
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 8 QV5SHITFL-NVADSTSTSEVYTTSTGSSSESAISTTSGIEISTTLEASTTDSQDS 66
0Y 80 ITDYL5FDJOKEGIFYPAKNTLPESGAIGASPSNPVEIRDTIGBVFIEPNTTCRPF 139
Db 67 STSTSDGTSOSTTIDBSNSTPSTDSGSLQSPDSSSASD5MPT----- 114
0Y 140 SSNPMAAVNKKIREGAIHAQULYLNHNHDVYGFMAKNSYV-----RCGALSTANT----- 189
Db 115 TVDPASTETPYD---FVLEMLTNV--ETIYSENEFPYITPIPKBPAGLTTAMQCQR 168
0Y 190 -----FVYSEMSOFLPMDNICLOTNAGKGAIVYGT5---NSP-ESNNCDPLFTNN 238
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 169 NDSOPEFLVLESNCL-----TEFGNCAIVASVSFNPMPTSVYPATGYEFLINW 218
0Y 239 ACCAGGAIFSPICSLTGNRGVIVEYNNRCFKNVETASSEASDGAIKVTTRLDVTGNRGR 298
Db 219 TMRAGCESASHIFPM-----NVVLPTTTTETRPPTTVSSSDGAGKGTGTCATGCTGTGTS 273
0Y 299 IFFSDNITKNNGAIYAPVYTLVDNGPYTFINNIANNKGAIYIDGINSNKT5ADRAIL 358
Db 274 --GG5AT1LTSTGDAVR5-----TTS6SGSGG5STG5GAGG5---GTTAGSGSGG5SGT 322
0Y 359 FENELVTNVTNA-NCTST5AMP5PRNAITVVASGCE--ILGACSSONLJFYPDIEV5NA 415
Db 323 GSDGVN5SGKTTALNDGDTGSG-----TATTPSHJLGDG5TSGSSDSN-----GSS 369
0Y 416 GVSVSFENKEADGTGSVEFSGAT-VNSADFHORNLOTKTPAPLTL5NGFLECIEDHAOLTVN 474
Db 370 GYTKRSSGSDPTSGSDSSD5GANGAF5ATAOP5TRTTRT5SLA----- 412
0Y 475 RFTOTRGCVAVSLONGAVL5CYKNKAGCN-----SASNA5TTLKHIGLNL5SILK5CAEI 526
Db 413 ----TV5PISAAEQIILIDQKRDVANNQLAGIMDG5ASNN5LNT5-----SLLNIQISL 462
0Y 527 P---LLW5EPTN5NNNYTADTAPAT5LSLDV5KSLIDVGN5PE5TD-----LTHAL55OP 579
Db 463 PAADLVEVAQ5LSL5TKLIPGVGNM55VDV5LKTLODNIATTN5EL5AD5MAK5VITKLANVN 522
0Y 580 M5ISEASDQNL5R5DM5E5GLN5PHY5G5GLW5GAK5TOD5EP-----5ASAT 629

```

```

Db      523 MTS-AOSLNSVLSLSDLALKSTV-----YPLGVSTSKSKDGTAAVIFGVIASGYT 5733
QY      630 ITDQ-KANRHRLLLTW-LPACVYSPKRRSPLINLTLMGNMLLATBSLKNSAELTPS 6687
Db      574 LVSPCTLSITVSTIYLTGDTBASY-----KOLDGDTYATDMLMAAI----- 6166
QY      688 DHPFGITG----GGLGMNMYQDRENNPGFHMRSSGYSAGMIA----GQTHFSLKFS 7388
Db      617 -----GLOGMATNGRIVQVEQBDLID-----KRLVSGINIAATMSGVDVQSGEVSYN 6655
QY      739 QTY-TKLNERYAKNNV---SSKNTSCOGEMLFSLQEGFLTK--LVGLYSYGDHNCNHF 7921
Db      666 DMVYTAAMVNTVDNSVTGTSQAKNTSEFENIPVSEVOYLLIESGTMTIKLHS----- 7166
QY      792 YTGQENLISQCTFFSQTMGCAVFDDLPMKPRGSHILLTAFLGLALGYSSLSHPTEGVAY 8511
Db      717 ---TONIVSRKLVYTAASGVTYITTCIN-GTQKEVEVDTNAINESYINDS-FTVYASD 7707
QY      852 PRSFSTKTLPLNLVLPV-GVYKGSFMANTORQAMTVELAY--OPVLVROEGIATOLLAS 9088
Db      771 GSSASTVKKLLQMPVIEVNLALFNQTTSLPVEFSNMGVSMRWVLSPDQIGIPAVSALS 8300
QY      909 KGIMFGSGSPSSRHA---MSRKISQOT 932
Db      831 QTVSISLSPFASVYTKDDQSLKEOT 857

```

RESULT 13
AIDA ECOIT

```

ID  AIDA_ECOLI STANDARD: PRI: 1286 AA.
AC  003155:
DT  01-JUN-1994 (Rel. 29, Created)
DT  01-JUN-1994 (Rel. 29, Last sequence update)
DT  01-JUN-1994 (Rel. 29, Last annotation update)
DE  ADHESIN AIDA-I PRECURSOR.
GN  AIDA-I.
OS  Escherichia coli.
OC  Plasmid p186.
OC  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC  Escherichia.
OX  NCBI_TaxID=562:
   [1]
RP  SEQUENCE FROM N.A., AND SEQUENCE OF 50-56.
RX  STRAIN=2787 (O126:H27);
RX  MEDLINE=92326638; PubMed=1625582;
RA  Benz I., Schmidt M.A.;
RT  "AIDA-I, the adhesin involved in diffuse adherence of the
RT  diarrhoeagenic Escherichia coli strain 2787 (O126:H27), is
RT  synthesized via a precursor molecule.";
RL  Mol. Microbiol. 6:1539-1546(1992).
CC  - FUNCTION: THIS IS AN ADHESION PROTEIN NECESSARY FOR THE DIFFUSE
CC  ADHERENCE OF DIARRHEA-CAUSING ENTEROPATHOGENIC ESCHERICHIA COLI
CC  TO EPITHELIAL CELLS.
CC  - SUBCELLULAR LOCATION: OUTER MEMBRANE.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL, X65022; CAA46156.1; -.
DR  PIR, S28634; S28634.
FM  Cell adhesion: Signal; Outer membrane: Plasmid.
FT  SIGNAL 1 49
FT CHAIN 50 2 ADHESIN AIDA-I.
FT PROPEP ? 1286
SQ  SEQUENCE 1286 AA; 132271 MW; B2A00F72AC05FB34 CRC64;

```

```
Db 1614 -DNVEGSNLTDAADITSTSVVWTAKGSTDASGVNDVMS-----KN 1654
QY 620 QDPPEASATITDPOKANRFRHTLLTWLPAGYVSPKHSRPLANTLMGNMLATES-- 677
Db 1655 AYDVATDASVNDIAKA-----LDAGYTNNEFLTSLNGTAELSALKOVSGS 1703
QY 678 -----LKMSAELTPSDHPFWGITGGGLGMVYQ--DPRE---NHGCFHM----- 716
Db 1704 QATTVREARVLSNRFSLADAP--KVNGLAFFNVAKDPREELGNTEYDMLAIR 1760
QY 717 -----RSSGYSA-----GMIAGOTHTFSLKESQYTK-LN---ERY 748
Db 1761 TIDSESGTMSUEVGLIARDLDDGAKACDNCVGTGYSQFGLKHMSPDNMNMNALRY 1820
QY 749 AKNNYSK-----NISCQCEMLFSLOEGELLTKLVGLSYGDHN 787
Db 1821 DVHNLDSRSIAFGNTKNTADTVKQOYLEFRSEGAKTPEPSEGLKVPYAGV-----K 1874
QY 788 CHHFTYQGENLTSGTFRSQTMGCAVFEDLPKMPGSGTHI-----LTPALGALGYSS- 841
Db 1875 LRH-----TLBGGVOERNAGD---FNLNNSGSETAVDISYGLKLDYAGKDGWSASA 1923
QY 842 -----LSHFE-----VGAYPRFST---KTPLIINVLPICVKS 873
Db 1924 TLEGPNLSYAKSORTASLAGAGSOFHNVDDGOKGGINSLTSGVKYS 1972

RESULT 10
190K RICRI STANDARD: PRT: 2249 AA.
AC P15921:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DE 01-OCT-1994 (Rel. 30, Last annotation update)
DE 190 KDA ANTIGEN PRECURSOR (CELL SURFACE ANTIGEN).
OS Rickettsia rickettsii.
OC Bacteria: Proteobacteria; alpha subdivision: Rickettsiales;
Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=783;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=R;
RX MEDLINE=90354033; PubMed=2117568;
RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;
RT "A protective protein antigen of Rickettsia rickettsii has tandemly
repeated, near-identical sequences."
RL Infect. Immun. 58:2760-2769(1990).
CC - FUNCTION: ELICITS PROTECTIVE IMMUNITY.
CC - PTM: GLYCOSYLATED (POSSIBLE).
CC - DISEASE: PROBABLY PLAYS AN IMPORTANT ROLE IN THE PATHOGENESIS OF
AND IMMUNITY TO ROCKY MOUNTAIN SPOTTED FEVER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M31227; AAA26380.1;
DR PIR: A41477; A41477
KM Antigen: Repeat; Signal: S-layer; Glycoprotein.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 2249 190 KDA ANTIGEN.
FT DOMAIN 212 1180 13 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 287 358 A (TYPE I).
FT REPEAT 359 430 B (TYPE II).
FT REPEAT 431 505 C (TYPE II).
FT REPEAT 506 577 D (TYPE I).
FT REPEAT 578 652 E (TYPE II).
FT REPEAT 653 724 F (TYPE I).
FT REPEAT G (TYPE II).
```

```
FT REPEAT 725 799 H (TYPE I).
FT REPEAT 800 874 I (TYPE I).
FT REPEAT 875 949 J (TYPE I).
FT REPEAT 950 1021 K (TYPE II).
FT REPEAT 1022 1093 L (TYPE II).
FT REPEAT 1094 1165 M (TYPE II).
FT REPEAT 1166 1180 TYPE I (INCOMPLETE).
SO SEQUENCE 2249 AA: 224333 MW: A9D6646C089DF087 CRC64;

Query Match 3.4%: Score 174; DB 1; Length 2249;
Best Local Similarity 21.2%: Pred. No. 0.0098;
Matches 164; Conservative 90; Mismatches 294; Indels 226; Gaps 34;

QY 10 IGNSLSGLAREPSPRIFLMPNSVDPPTKESLSKIS-----LTGDTNHLNLTCLD 59
Db 212 IGNTNALATVWVGATVGLGAVIKATTKLTNNASVLTNNANAVLTGALDNTGG--D 269
QY 60 NLRYL---ATLOKTPNEGAAVTTIDVLSFDTQKEGIVFAKNLT---PESGALGYASP 113
Db 270 NVGVLTNGALSQVTTGDTGNTNSLTATISVGAGTATLGGAIVKATTKLTLDASAVKFTNP 329
QY 114 NSPTVEIRD---IGPVIFE-NNTCCRPFTSSNPNAAVNRIR---EGGAIHAQNLVI 163
Db 330 VVVTGALDNTGNANNGIYVTFGTGNSVTGNGCNTNALATVWVGAGLLQVGGVVANTINL 389
QY 164 NNINIDVGFMAKNFSYVRGGAISTANTFVSENOGCFLEMDN-----ICIQ 208
Db 390 TDMASAVTF-TNPVVVTAIDNTGNA-----NNGIVFTGNSVTGDTGNTNALATVWVG 443
QY 209 TNTACKGATVAGTSNSPESNNCDLFFINNACAGALFES-----PICSLT 254
Db 444 AGTATLGGAIVKATTKLTNNASVLTNNANAVLTGALDNTGGDNVGLNGLALSQVT 503
QY 255 GNRGNIVFYNRCEKNVETASSEASDGA-IKVTT-----RLDVT 293
Db 504 CNIG-----NTNSLATISVGAGTATLGGAIVKATTKLTLDASAVKFTNPVVVGTALDNT 558
QY 294 G--NGRIFFSDNT-----TKNNGAIYAPVTVLVNDGPIYFINN 331
Db 559 GNANNGIVFTGNSVTGDTGNTNSLATISVGAGTATLGGAIVKATTKLTNNASVLTTL 618
QY 332 JANNNGAIYIDGTGNSKISADRHAIIFENENIVTWTNANGSTSANPRRNATVASS 391
Db 619 NNN-----AVLTGALDNTGGDNVGLNGLALSQVTGDTGNTNS-----LATISGA 666
QY 392 GEILLAGSSONLIFYDEIEVSNAGVSSEFKEA-----DOTGS---VVFSGATVNS 440
Db 667 GATATLGCA---VIKATTKLTNNASAVKFTNPVVVGTALDNTGANNNGIYVTFGTGNSVT 722
QY 441 ADFHORNLT-----QTPPAVLISNGFLCIEDHQAULTVRF 476
Db 723 GDIGTNNLATVWVGATVGLGAVIKATTKLTNNASVLTNNANAVLTGALDNTGG 776
QY 477 TOTG---GVVSLGNGAVLSCKKNGASNASASTLKHIGLINSILKSGAEPLLWVE 532
Db 777 NTTGGDNVGLNL-NGA-LSQVTTGDTGNTNSLATIS---VAGATVTL--GGAVIKATTK 829
QY 533 PTN-----NSN-----NYT-ADTAATFSLSVVKLSLIDDYGS----- 564
Db 830 LTNASVLTLTNNANAVLTGAVDNTGGDNVGLNGLALSQVTGDTGNTNSLATISVAG 889
QY 565 -----PIESTDLTALSSQPMLSISEAS-----DNOLRBDMDPFSGLN 602
Db 890 TATLGGAIVKATTKLTNNAS---VLTLTNNANAVLTGALDNTGGDNVGLNLN 940

RESULT 11
VAC3_HELPY STANDARD: PRT: 1310 AA.
ID VAC3_HELPY
AC Q48253:
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
```

Db 1050 -LSNSTSNIGTKANNGTIVYLGAENIGNICSDSDTPVAVSFRTGSDSAGIQCNIQSVI 1108
QY 442 DFIHRLNLOTKTPAPLTLNSGFLCIEDHAQLTVNRFOTGVVSLONG---AVLSQYKNG 497
Db 1109 DFCYVNLGI-VNSNIIIGCGTALNGKIDLVNTLTFMSGTNGNNTSIEITLLANGN 1167
QY 498 AG-----NSASNASTLK-----HIGLNSLSIISKCAEI-----PLWEP 533
Db 1168 IGHVILEGAQVNTLTGTTTIFKIVODNANANFSGTQITLLTGGRFNGTLGSPFAVTG 1227
QY 534 TNSNNTYADTAAT-----FSLDVKLSLIDYGNPY-----ESTDLHALSQPM 580
Db 1228 SNREYVNSLIRAAADQVYITRTNNAENVNTDIANSFPGACPVQDQNTTVNANATNAY 1287
QY 581 LSISEASDNLRSDDMDFSGLVNPIHYGQGLTWGMATQDPEPASATITDPO----- 634
Db 1288 NNULLAKNS---ANSANFVGAIVTD-----TSAITVQLDLAKD 1324
QY 635 -KANFHTLLTLTLPACIVSPKHSPLANTLGMNMLATESLKNABELTPSDHFWG 693
Db 1325 IOAQLGNRLGALRYLG---TPETAEMAGPEAGAISAAVAGDEAIDNVA-----YG 1372
QY 694 ITGGGLGMVYQDPEHNPGRHMRSSGVSAQN--IAGQTHFSLKFSQTYTKL-NERYAK 750
Db 1373 IWAPEYTDAAHQSKKGLAGYKAKTGVVIGLDTLANDNLMGAIGITKIDIKQDYKK 1432
QY 751 NNVS-----KNYSCGEMLEFSLQDFLLTKLVLYSYGDHNCIHFTYG 795
Db 1433 GDKTVNGFSFSLYGAQQLVKRFEAGSIFELNQ-----VKNSQRYFEDAN 1480
QY 796 ENLJTSQ---GTFRTQMG--AVFDLPKPRGSHILITAPFLGALGYSLSHTEVGA 850
Db 1481 GMSKQIAGHVDNMTFGNLTGVGDYN---AMQGLVTPMAGLSYKSSDENYKEGT 1536
QY 851 -----YPRFSTKPLINLVPIGVKGSFMANATORPQMTLAYQPLLYR----- 896
Db 1537 TVANKOVNSKFSRDITL---VGAKVAGSTNMT-----DLAVYPEVHAFFVHKYTC 1585
QY 897 -----OEPGIATOLLASKGIMFGSGSPSSHHAMSYKISQOTPLSWLT 939
Db 1586 RLSTQSVLDGQVTPCINQDRTKTSYVLGL---SASIRSDAKMEYGIQYDAQISSKYT 1642
QY 940 LH 941
Db 1643 AH 1644

RESULT 8
OMP_RICJA STANDARD: PRT: 1656 AA.
ID OMP_RICJA 006653;
AC 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCAS) (ROMP B)
DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB): 32 KDA BETA PEPTIDE).
GN OMPB.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsia.
OX NCBI_TaxID=35790;
OY 111
RN SEQUENCE FROM N.A.
RP STRAIN-YH:
RC uchiyama T.;
RT Sequencing of the gene encoding the protein romp B of Rickettsia
RT japonica.
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY

CC SIMILARITY).
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AB003601; BAA20138.1;
DR Antigen; S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.
KW CHAIN 1 1338
FT CHAIN 1339 1656 32 KDA BETA PEPTIDE.
FT DOMAIN 528 533 POLY-GLY.
SQ SEQUENCE 1656 AA: 168097 MW: 3132A69C9DD5999F CRC64;
Query Match 3.7%: Score 189; DB 1; Length 1656;
Best Local Similarity 19.2%: Pred. No. 0.00076;
Matches 218; Conservative 129; Mismatches 428; Indels 358; Gaps 49;
QY 89 TOKGICYFAKMLTPESGAGIYASPSPTVEIRDTGIVTEENNTCCRP-----TSS 141
Db 592 TDQGVVAVASSLT-----NAQTLTISGIG-IIGANNNTLQGFNIGSSKTLN 638
QY 142 NNAAVNKI---REGCAIHAONLY-----IHNHDV-----VGFMK 174
Db 639 GGNVAINELVGNNGSVOPAHNTYLTITRTNAGGAKIIPNPVNNNTTLAAGTLGSA 698
QY 175 N-FSVYRGGAISTANTFYVSENOGSEFLMDNICIOTNAGK-----GGAIVAGT----- 222
Db 699 NPLAEINFGSKCARADYVLNVEGVNLVATNITTDANVGSFVNAGKNIVSGVGGQ 758
QY 223 SNSEFS---NQCDFEFINNACCAGATFS----- 248
Db 759 GKKFTVALDNGTIVKFLGNATFNNTTIAANSTLOISGNTADFIASADGTGIVEYNT 818
QY 249 -----PICSL-----TGNRGN-----IYFYNRC----- 267
Db 819 GPINVLTKQAVPVNAKQITVSGPNVYVNEIGNAGNTHGAMTDITAFENSSLCALVFL 878
QY 268 -----FKNV-----ETASSEASDG-----GAIKVTRLDYTGNGRI-----F 300
Db 879 PSQIFENDAGNTIPLTIKSTVGNETAEGFSPSVIYSGVDVYIADGQVIGDQNNIVGLQL 938
QY 301 FSDNTITKNGAIYAPVTVLVNGPTFYFINNANKGAIYIDGT--SNKISADRHAI 358
Db 939 GSDNGIYVATLTYAGIGINNOCGTVLSGVPTNPFGVYLGIGIGASKEKQVFTTD 998
QY 359 FNE--NIV-TVNTNANGSTSANPPRRAIIVASSGCEILGA--OSSONLIYDPI----- 410
Db 999 YNNGLIINTNTTINDGVTVTTGGIAGGIAGTDIDGKITLLSYVGNANVPRADGIFSNS 1058
QY 411 -----EVSNAVSVSFKREADQTS-----SVFSGATVNSA-----DFHQ 445
Db 1059 TSMIVTTKANNGVTVYLGNAFVGNIGDSDTPVAVSFRTGSDN-NGAGLKNIVSOVIDFGT 1117
QY 446 RNLQTRPAPLTLNSGFLCIEDHAQLTVNRFOTGVVSLONGAVLS----- 492
Db 1118 YNLGI-VNSNVILGSGTIAINGKIDLTNTLTFAGGTSTWGNNTSIEITLLANGNIGHI 1176
QY 493 CYKNAGNSASNASTLKHL-----GLNLSLSIISKCAEI-----PLWEPNNS 537
Db 1177 VIAEGQVANAATTGTTTIVODNANANFSGTQITLLTGGRFNGTLGSPFAVTGSMNF 1236
QY 538 NNTYADTAAT-----FSLDVKLSLIDYGNPY-----ESTDLHALSQPMLSTS 584
Db 1237 VNYGLIRANQDVYITRTNNAENVNTDIANSFPGACPVQDQNTTVNANATNAYNNLL 1296

QY	156	IIAHQNIYIHNHNDYGFMMKPFYVAGAGIASTANTFVVEENOSCFLEPMONIC1QNTA--G	213
Db	115	IFAKENSTLNTLTVI-FSGNVAGVGGIAYSSGT---NDTGAVDLRYTNMFRNNINDG	170
QY	214	KGAIYAGTISNFSFNCCDLFFIINNACCAGAIIFSPICSLTGNGNIYFYNNRCFKVET	273
Db	171	KGAIY-----TINNYYIL-----SDVIFDNNQAV---TS	197
QY	274	ASSEASDGAIAKVTFLDYLTGNR-----GRIFESDNIRKNYGAIV-----AP	316
Db	198	TTSISDGDGAI-----DVTDNNSDKHPSGCTTYNNATFTNNTAEIGYGAITYNSYAP	251
QY	317	VVTLVYDNGPTPYFINN---IANNKGAIYIDGTSN-----SKISADRAH11FENI	363
Db	252	YLLDISVDSYQNGGVLVDENNNSAAGYGDGPPSSAAGGFMYLGLSEVTFD---IADGKTL	308
QY	364	VTVNTVANGTSTANPPRRRAITVASSSEIILLGAGSSGNLIFYPIEVSNAAGSVSRNK	423
Db	309	VIGTENDGAGVDSI---AGTLTKTKGSDDLVLMNDND---FTGEMOIENGEVYLCRSN	362
QY	424	EADDTGSVFS-----GATVNSADFHORNL-----QTKTPAPLTLNSGFLCIEDH	468
Db	363	SLMNVGTHQDDPODCYGLTGISIDQYONQANELMVGSTQGFVHALGFGQGLINTDAG	422
QY	469	AQLTVNFTQTCGVVSLGNCGAVLSCYKNCAGNSASNASITLKHLGLNLSILKSGAEIPL	528
Db	423	GNVTYVNGSPAGIIEGAGQ---LTIQNSSYVLAGASQSMALTG---DIYVDDGAVLSTL	474
QY	529	LWNEPTNNSNNYADPAATFSLSDVATLSLIDYG-----NSPYESTDLTHALSQPMLS	582
Db	475	-----BGDAADLTALODDPQSTIYLGVLSDLSDFSTWOS-----GTYNGLLE	517
QY	583	ISEASDNQLRSDP-MDFSGLVNPHYWGQ	610
Db	518	VSGSGTVIGSDPVVPLAGDNLHIGDG	546
RESULT 4			
OMP	OMP	RICPR	STANDARD: PRT: 1643 AA.
AC	053020:	09ZCM0:	
DT	30-MAY-2000	(Rel. 39, Created)	
DT	30-MAY-2000	(Rel. 39, Last annotation update)	
DT	30-MAY-2000	(Rel. 39, Last annotation update)	
DE	OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)		
DE	(SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCAS) (ROMP B)		
DE	(CONFINE: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)		
DE	(120 KDA OUTER MEMBRANE PROTEIN OMPB): 32 KDA BETA PEPTIDE).		
GN	OMP B OR SPA OR SPA OR RP704.		
OS	Rickettsia prowazekii.		
OC	Bacteria: Proteobacteria: alpha subdivision: Rickettsiales:		
OC	Rickettsiaceae: Rickettsiidae: Rickettsia.		
OX	NCBI_TaxID:782;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.		
RC	STRAIN-BREINL:		
RC	MEDLINE-91045972; PubMed-2122457;		
RA	Carl M., Dobson M.E., Chung W.M., Dasch G.A.:		
RT	"Characterization of the gene encoding the protective paracrystalline		
RT	surface-layer protein of Rickettsia prowazekii: presence of a		
RL	truncated identical homolog in Rickettsia typhi."		
RL	Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).		
RP	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BREINL:		
RC	Morion C.G., Yu X.J., Walker D.H.:		
RT	"Sequence analysis of ompB of Rickettsia prowazekii."		
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-MADRID E:		
RC	MEDLINE-99039499; PubMed-9821893;		
RA	Andersson S.G.E., ZomorodiPour A., Andersson J.O.,		

RA Sticheleit-Ponten T., Alsmark U.C.M., Podowski R.M., Naestlund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.:
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria."
RL Nature 396:133-140(1998).
RN 141
RN PARTIAL SEQUENCE.
RP STRAIN=BREINL;
RX MEDLINE=92114896; Pubmed=1370573;
RA Ching W.W., Carl M., Dasch G.A.:
RT "Mapping of monoclonal antibody binding sites on CNBR fragments of
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT prowazekii."
RL Mol. Immunol. 29:95-105(1992).
RN 151
RN IDENTIFICATION OF CLEAVAGE SITE
RP MEDLINE=92104668; Pubmed=1729180;
RX Hackstadt T., Messer R., Cieplak W. Jr., Peacock M.G.:
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
RT membrane protein of rickettsiae: identification of an avirulent mutant
RT deficient in processing."
RL Infect. Immun. 60:159-165(1992).
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as their content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb.slb.ch/announce>
CC or send an email to license@isb.slb.ch).
CC
CC EMBL: M37647; AAA26390.1; ALT. INIT.
DR EMBL: AF161079; AAD42234.1; -
KW EMBL: AJ255273; CAA15140.1; -
KW Antigen: S-layer; Cell wall.
FT CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1329 1643 32 KDA BETA PEPTIDE.
FT VARIANT 257 257 Y -> A (IN STRAIN BREINL).
FT VARIANT 1010 1010 Y -> D (IN STRAIN BREINL).
FT VARIANT 1450 1450 A -> S (IN STRAIN BREINL).
FT VARIANT 178 179 AA -> VC (IN REF. 1).
FT CONFLICT 191 201 TTQEWPLTGLA -> INRSSTSYHLVS (IN REF. 1).
FT CONFLICT 212 212 Q -> I (IN REF. 1).
FT CONFLICT 313 313 Q -> L (IN REF. 1).
FT CONFLICT 1104 1104 D -> G (IN REF. 2).
FT CONFLICT 1123 1123 T -> S (IN REF. 2).
SO SEQUENCE 1643 AA: 169854 MW: 735FDF32BE346CC CRC64;

```

Query Match 193: 3.8%; Score 193; DB 1; Length 1643;
Best Local Similarity 19.1%; Pred. No. 0.00043;
Matches 192; Conservative 133; Mismatches 434; Indels 244; Gaps 37

QY 41 SKKISLTGDT-----NHULTNCELDNLRYTLLAIDLOKTTNECAAVYITTDYLSFFDQREG 93
DB 633 SKTILNAGVAINELVIEENGSVQLAHNTYLLTKTTINAAOQDIYAADPLNNTITLAAIG 692
QY 94 IYFAKNLTPESGAIGAYASPNSPVEIRDTIGPIFENNCTCRPFTSSPNNAVANKIREG 153
DB 693 TMLGAENPLSTIHFKAKANADSI-LNQGCKVNLAYNN-----ITTDANNGSLFHRSG 746
QY 154 GAI-----HAONLYIHNHHGVGFEMKFFSVRGCAISTANTFVASENOSC----- 198
DB 747 GTSIYSGTGGGQGHKLNLLIDNCTTVAFLDGTTFENGSTIEGKSLILQISNNYTTDHYE 806
QY 199 -----FLEMDNICIOTNTAGKCAIYAGTSNFSFNCDLFF--INNACGAGCAIF 247

```

```

OY 181 GGAISTANTFVASENOCFLFMDNICIOTNTAGKGAIIYAGTSNFSFNCDLFFINAC 240
DB 180 GCAISTANTFVASENOCFLFMDNICIOTNTAGKGAIIYAGTSNFSFNCDLFFINAC 239
OY 241 CAGATFSPICSLTGNRCNIVFYNNRCKNVEYASSASDGAIKVTRLDVYTGRCRIF 300
DB 240 CAGATFSPICSLTGNRCNIVFYNNRCKNVEYASSASDGAIKVTRLDVYTGRCRIF 299
OY 301 FSDNITKNGGAIYAPVTVLVNCPFTYFNNIANNKGAITYIDGTSNKSISADRAHAI 360
DB 300 FSDNITKNGGAIYAPVTVLVNCPFTYFNNIANNKGAITYIDGTSNKSISADRAHAI 359
OY 361 ENIVTNTNANGTSTANPPRRNATTVASSGEILLGAGSSONLIFYPIEYNAGVYS 420
DB 360 ENIVTNTNANGTSTANPPRRNATTVASSGEILLGAGSSONLIFYPIEYNAGVYS 419
OY 421 FNKEADOTGSVFGATVNSADFHORNLOTPAPLTLISNGCLTEDHAOLTVNFTOTG 480
DB 420 FNKEADOTGSVFGATVNSADFHORNLOTPAPLTLISNGCLTEDHAOLTVNFTOTG 479
OY 481 GVSLSGNCVAVLSCKYKNGAGNSASNASITLKHIGLNLSSILKSGAEIPLLWEPITNSNY 540
DB 480 GVSLSGNCVAVLSCKYKNGAGNSASNASITLKHIGLNLSSILKSGAEIPLLWEPITNSNY 539
OY 541 TADTATATSLSDVKLSLIDDYGNSPYESTDLTHALSSQPMLSISASDQNLASDMDTSG 600
DB 540 TADTATATSLSDVKLSLIDDYGNSPYESTDLTHALSSQPMLSISASDQNLASDMDTSG 599
OY 601 LNVPRYHOGGLTWGMWAKTODEPAPASATITDPOKANRPHRTLLTWLPAGVPSPKHRS 660
DB 600 LNVPRYHOGGLTWGMWAKTODEPAPASATITDPOKANRPHRTLLTWLPAGVPSPKHRS 659
OY 661 PLIANTLMGNMLLATESLKSNAELTPSDHPFWGITGGLGMVYODPRENHPDPMHRSRG 720
DB 660 PLIANTLMGNMLLATESLKSNAELTPSDHPFWGITGGLGMVYODPRENHPDPMHRSRG 719
OY 721 YSAGMIAGOTHTFSLKFSOTTKLNERAKANNVSSKNVSCGEMFLSLOEGLLTKLVGL 780
DB 720 YSAGMIAGOTHTFSLKFSOTTKLNERAKANNVSSKNVSCGEMFLSLOEGLLTKLVGL 779
OY 781 YSYGDNCHHPTGEBNLISGCTPRFSGOTMGAVFPDLPKPKGSTRHILAPLALGATIS 840
DB 780 YSYGDNCHHPTGEBNLISGCTPRFSGOTMGAVFPDLPKPKGSTRHILAPLALGATIS 839
OY 841 SLSHTEVGAYPRSFSTKPLINVLPIGVKGSFNNATOPQAWTVELAYQVLYROEBG 900
DB 840 SLSHTEVGAYPRSFSTKPLINVLPIGVKGSFNNATOPQAWTVELAYQVLYROEBG 899
OY 901 IATQILASGIVFGSGSPSRHAMSTKISQOTOPLSMLTLHFQYHGFYSSTFCNTLNGE 960
DB 900 IATQILASGIVFGSGSPSRHAMSTKISQOTOPLSMLTLHFQYHGFYSSTFCNTLNGE 959
OY 961 IALRF 965
DB 960 IALRF 964

```

```

RESULT 2
OMPF_CHLTR STANDARD: PRT: 1034 AA.
AC P38008: 084878:
DT 01-OCT-1994 (Rel. 30, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PUTATIVE OUTER MEMBRANE PROTEIN F. PRECURSOR.
GN PMPF OR CT870.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxId=813;
OX 11
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/CX;
PX MEDLINE=99000809; PubMed=97844136;

```

```

RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.:
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
RN [2].
RP SEQUENCE OF 26-35.
RC STRAIN=L2/434/B0;
RA Bini L., Santucci A., Magi B., Marzocchi B., Sanchez-Campillo M.,
RA Comanducci M., Christlanen G., Birkelund S., Vitreou E., Ratti G.,
RA Pallini V.:
RL Submitted (SEP-1994) to the SWISS-PROT data bank.
CC -I- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001360; AAC68468.1; -
KW Outer membrane; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1034 PUTATIVE OUTER MEMBRANE PROTEIN F.
FT CONFLICT 34 35 RR -> FH (IN REF. 2).
FT FT
SQ SEQUENCE 1034 AA: 112392 MW: 445FAC35D463A67 CRC64:

```

```

Query Match 17.0% Score 866.5; DB 1: Length 1034;
Best Local Similarity 28.1%; Pred. No. 76-46;
Matches 293; Conservative 149; Mismatches 406; Indels 193; Gaps 36;

```

```

OY 55 NCYLDNLRVYIALILOKTNENGAIVITDYLSPFOKEGIVFAKRLPESGAI----- 108
DB 57 NIVLSNLO-----SNGTGACTISG-----NTOIFNSVNTVTTDSCGAEPMWYTS 102
OY 109 -----GYASNPSPVEIRDTGPIVFNENNTCCRFSTSNPAV-----NK 149
DB 103 FTASDNANLFCNNCTHNKGGAIKRG-GPIRLNNDVLFYNNISAGAYVGTGDNE 161
OY 150 IREGAIIHONLIVHNIDVYGFKNFSYVKGAIANTFVASENOCFLFMDNICI-- 207
DB 162 KNRGALVATTTTLGNR-TLAFINNMSSGDCGAIADTOISITDVGILFFENHRTLNH 220
OY 208 -----QTNAGKGAIIYAGTS-NSFESNCDLFFINNACGAGIIFSPICSLTGNRCNIV 262
DB 221 IPTYOENNAKRGALICSRDLDCISNSGPIYENYOGKGAISATKCVLDNNKEHIF 280
OY 263 YNNRCFKNVEYASSASDGAIKVTRLDVYTGRCRIFFSNITNKGAIYAPVTVLV 322
DB 281 SNSSILG--WSQSSASNGAIQITOGFTLRNNKSIVFDSMTATHAGALINCYIDRD 338
OY 323 NGPTYFINNIANNKGAIIYIDGTSNKSISADRA-----IIFENYIYNTVNTANG 372
DB 339 NCPVYFLNNSA--AWGAAP-----NLSPRSATNYIHTGTGDIVPNNNVFTLDGNLL 389
OY 373 TSTSNPPRRNATTVASSGEILLGAGSSONLIFYPIE-----VSNAGS-----V 419
DB 390 GKRLFLHINNNEITPTLS-----LGAKKDRIRYFDLQWEMKENTSNPSPISRTI 445
OY 420 SFNKEADOTGSVFGATVNSADFHORNLOTV-----TPAPLTLISNGCLTEDHAOLTVN 474
DB 446 TVNPETEFGAVVFS--YNOMSSDIRLMDKEHNYIKAPFTLFLGTLAIEDDALEIF 502
OY 475 R--FTQT-GGVVSLNGAIVLSCKYKNGAGNSASNASITLKHIGLNLSSILKSGAEIPLMW 531
DB 503 NIPFTONPTSLALGSGALLTVGKHG-----KLNTNLGVLLPILIKESKSPCLIRV 554
OY 532 EPTN-NSNNYTDAT--AATSL-----DVKLSLIDDYGNSPYESTDLTHALSSQPMLSI 583

```

Mon May 7 09:06:31 2001

us-09-677-752-2.rag

Page 24


```
OY 559 DDYGNSPYESTDLTHALSSQPMLSISEASDNQLRSDMDPFGSLNV-----PHYMGQL 611
D 525 dssgnf-yeshntingaf-tqplvftaat-----aasdiydalltspvqtpdephgyg9h 578
OY 612 WTWGMAKTODPEPASSATITDPOKANRFRHRTLTLTLWLAGVVPSPKHSPLIANTLWGMN 671
D 579 weatwa-----dstaksgt-----mlwtvtgynpnpertiasvpdsiwaaf 620
OY 672 LLATSLKNSAELTPSDPFW---GITGGGLGMVYODPRENHPCFHNRSQGSACMIAG 728
D 621 ---tdirtlqimtsqansiyqrglwasgtanftfhkksyngaftrhksyiyvggsae 677
OY 729 --QTHFSLKFSQTYTKLNERYAKNNVSSKNYSCQ-----GEMLFSLQ 769
D 678 dtsenifsvafcgltfgkdkdlftiven-tshnylaslyqhraflgglpmpsfgsitdmjk 736
OY 770 EGFLLTKIAGLYSYGDHNCNHFYT-----QGENLTSQCTFRSQTGAVFPDLPMK-PCG 823
D 737 dipllnaqlsysytkndmdtrlytsypaagawtinsgale---lqgsialyipeaplf 793
OY 824 STHILTAFLGALGIYSSLSHFTFEGAVPRSPSTKPTPLINLVPIGVGCFMNATQRPQA 883
D 794 ggyf---pfllkqgavysyqgnfksgaerafd-dgdlvncsipvgirlekisedek-nn 848
OY 884 WTVELAYQPVLYROEPGIATOLLASKGIW 912
D 849 feislaiygdvytrknprrtslmsvasw 877

RESULT 22
190240
ID Y90240 standard: Protein: 930 AA.
XX
AC Y90240:
XX
DT 29-AUG-2000 (first entry)
XX
DE Chlamydia antigen CPN100639.
XX
KW Chlamydia antigen: diagnosis; infection; community acquired pneumonia;
KW therapy: upper respiratory tract disease; bronchitis; sinusitis;
KW asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma.
XX
OS Chlamydia pneumoniae.
XX
PN WO200032794-A2.
XX
PD 08-JUN-2000.
XX
PF 01-DEC-1999; 99WO-CA01147.
XX
PR 01-DEC-1998; 98US-0110339.
XX
PR 01-DEC-1998; 98US-0110340.
XX
PR 01-DEC-1998; 98US-0110427.
XX
PR 01-DEC-1998; 98US-0110428.
XX
PR 01-DEC-1998; 98US-0110438.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Murdin AD, Oomen RP, Wang J:
XX
DR WPT: 2000-412339/35.
XX
DR N-PSDB: A30853, A30854.
XX
PT Nucleic acids encoding polypeptide antigens from Chlamydia useful for
PT preventing, diagnosing and treating diseases such as community acquired
PT pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset
PT asthma.
XX
XX Claim 16; Fig 7; 174pp; English.
XX
PS This sequence is a Chlamydia antigen of the invention, designated
CC CPN100639. The nucleic acids (and their complementary sequences) may be
```

```
CC used as diagnostic agents for detecting the presence of nucleic acids
CC encoding Chlamydia antigens in samples according to standard methods,
CC and therefore, for diagnosing Chlamydia infections. For example, they may
CC be used as primers and probes for diagnostic polymerase chain reaction
CC (PCR) assays. Antisense sequences may be used to down regulate
CC expression of the proteins and may be used to treat infections. The
CC nucleic acids may also be used to produce the protein antigens they
CC encode according to standard recombinant DNA methodologies. The
CC proteins may then be used as antigens for the production of antibodies
CC (i.e. as vaccines) for preventing infection by Chlamydia. The
CC antibodies may also be used as diagnostic reagents for detecting
CC infections. Chlamydia is a pathogen implicated in the development of
CC (for example) community acquired pneumonia, upper respiratory tract
CC disease (especially bronchitis and sinusitis, asthmatic bronchitis,
CC adult-onset asthma and acute exacerbations of asthma in adults.
XX
SQ Sequence 930 AA:

Query Match 12.1%, Score 615, DB 21: Length 930:
Best Local Similarity 27.0%, Pred. NO. 4.2e-37:
Matches 251; Conservative 140; Mismatches 374; Indels 164; Gaps 40:

OY 47 TGDTHNLNICYLDNRRLALILQKTPNEGAAYT---INDLYSFQTOKEGIYFAKNTPE 103
D 50 tada-nglnhyvysgnvyl-----ndagkyraltgcctfctgdlftgkygslfn-lvd 102
OY 104 SCGATGYASPNSPFVEIRDT---IGPVLFFENNTCCRPFTSSNPAVANKIREGCAIHAQ 159
D 103 agsnagaastadkallftgfnslftlaapgtlvagskstlssagalnldnglillsq 162
OY 160 NLRYNNHNDVGFMAKFSYVRGCAISTANTFYVSENOSCFEMDICYOTNTAGK-GGAI 218
D 163 nvsneann-----ggalt-klslsgnts-----slftsnakkiyga1 203
OY 219 YAGTNSPFESNNCDLFFINN-ACCAGAI-FSPICSLTGNRCNIYFVNNRCKNVEETSS 276
D 204 yssaaaslsqnlqglvtimnkgelvggalgfaasslqgn-sllfsgn-----tad 255
OY 277 EASDGAIRV-----TTRLVDVTGNRGRIEFSQNTIRKNGCAIYADVTLVDNGPTVFINN 331
D 256 aagkgalcyektgelgcpulltllsgnks-ltfaensvlggacahgldtsaagplffsn 314
OY 332 IANN-----KGAIIYDGTNSKISADRAIIFRENIVVTWNTANGTSTSANPRRNATV 387
D 315 rcgnlaagkgagalaadsqslsasnqgdlftlgnl-----tstsapstrtnaiyl 366
OY 388 ASSSGEILGACSSONLIRYDPIEVSNAGVS-----VSFKKADQGSVVFSGATVNS 440
D 367 gesaktlnlraagqslfyfdpaaetlgaadvllnqpdnspldygsltlvfgskl-s 425
OY 441 ADFHQR--NLQKTPAPITLSNGFLCIEDHAQLYVNRFTQTVGVSILGAVLSCYKNGA 498
D 426 adeaakaadnftslkqplalaagtlalkgnvelvngftlgegstll-----mqpqt 477
OY 499 GNSASNASTITLKHIGLIGLSSILKSGAEPLLVPEPPNNSNNTADTAATFSISDVKLSLI 558
D 478 klkadteaistlcklvldlsal-----egnkvsvlaetaqaanktilts-plvfq 524
OY 559 DDYGNSPYESTDLTHALSSQPMLSISEASDNQLRSDMDPFGSLNV-----PHYMGQL 611
D 525 dssgnf-yeshntingaf-tqplvftaat-----aasdiydalltspvqtpdephgyg9h 578
OY 612 WTWGMAKTODPEPASSATITDPOKANRFRHRTLTLTLWLAGVVPSPKHSPLIANTLWGMN 671
D 579 weatwa-----dstaksgt-----mlwtvtgynpnpertiasvpdsiwaaf 620
OY 672 LLATSLKNSAELTPSDPFW---GITGGGLGMVYODPRENHPCFHNRSQGSACMIAG 728
D 621 ---tdirtlqimtsqansiyqrglwasgtanftfhkksyngaftrhksyiyvggsae 677
OY 729 --QTHFSLKFSQTYTKLNERYAKNNVSSKNYSCQ-----GEMLFSLQ 769
```

```

QY 99 NITPESGGAICVYASPNSTVEIRDTIGVIFENNTCCPFTSSND-NAAV----- 147
DB 122 --llægfkelsfsmcnslavl-----paatlknysgfpclttspsngltyskcll11n 174
QY 148 -----NKIR-EGGAHQAQNLVYINHHHDVGVCFMKNSYVRGGAISIRANFEVSENOSC 198
DB 175 nektsfysnlvsgdgaidskclvrglsklcvfqlentagdgagcvvstfsmancap 234
QY 199 FLFMDNICIQTNTAG-KCGAIVA-----GTSNSFE-----SNDCDLEFINNACCAG 243
DB 235 lafvannv-----agvrgglaavdgqgqvssstledpvsfmrnlaveldgnavrg 288
QY 244 GAIFSPICSLTGNRCNIVFYNN--RCPKN-----VETASEASDCGAI 284
DB 289 gglys-----ygnvalllnngkclflinnvaspylaakqrlsgaaslnmnydggal 340
QY 285 KVTTRLDVTCN-----RGRIFESDNITKNVGAIVAPVTLVDNGPTFENNINANN 335
DB 341 fckngagagsgnsvsgsfdegvgnvfssnvaagkgaalyakklsvanccpvgfltnland 400
QY 336 KCGAIIYIDGTSNKSISADRHAIIFENI-----VTNVTNANGTSTSANPRRRAITVASSS 391
DB 401 -ggalylgsgelslsadylgdlifglnlkrakenaadvnglvs-----qalsmgsgq 454
QY 392 GEILGAGSSONLIFYDPIEVS-----AGVSVSPKKEADOTGSVHSCATVNSADPH 444
DB 455 kltllrakqghllfndplemangnmpaqsskllkldngqytdgdivfan---gsstly 511
QY 445 QNNLQTKTAPVLTLSNGFICIEDHQLTVNRFOTGGVAVSLCNGALVSCYKXGAGNS--- 501
DB 512 qn-----vllæggrlvtlrekaklsvnsjsqvg-----slyme-advstcdf 551
QY 502 -----ASNASTLTAKHIGLNLSSILKSGAEIRPLWVEPTN--NSNNTYADTAATES 549
DB 552 vcpqppqpqaanqltlnhlsllsllanna-----venpnltpadqsdhavlsgtclt 606
QY 550 LSDVKLS---LIDDYGNSPYESTD-----LTHAISQDPMISISEASDNQLRSDDM 596
DB 607 aegvltlsqglfledldtdaydrydwlgnsqklnvklqlgkx--panapsdclt----- 659
QY 597 DFGSLNVPHYGNQGLTWGMMAKTODEPPASATITPOKANRHRHRLTLTWLPAGVPS 656
DB 660 ---gmepkygygsvklaw-----dntamngpytlkatwtctgynpmp 701
QY 657 KHRSPLIATLMGNMLLATESLKSNAELTPSDHPFV-GITCGGLGMVYQDPRNHPGFH 715
DB 702 ervaslvpmnlvgs--lldrshalsalqaavdgsyrcrjlwsgvsnffyhdcldalqgyr 760
QY 716 MRSSGYSAQMIAGQTHTFSLKFSQYTKLNERAKNNV--SSKNYSCQCEMLPSTDEGFL 773
DB 761 ylsqgyslg---ansyfgsamfglatlevfgr--skdyvvcvrsnhhacisvylstqda-- 814
QY 774 LTLVLGLVSGDHNCHHFTYOG-ENLTSGTEFRSOT-----MCGAVFEVDPKMPFGS 824
DB 815 ---lcsyylfgdaflraasyvgfngqhmksyctfaeesdvrvdmncldgagjlpivltps 871
QY 825 THILT--APFLGALGYSSLSHTEVGAVPRSFSTKPLINVLVPIGVKGSFNNATORPO 882
DB 872 klylneirpivgdegfesfadshefleeqdgatks--ghllnlavpvgkldrcsstlhp 929
QY 883 AMTVELAVOPVLYROEBGATATOLLASKGIWFGSGSPSSRRAMSKYKISQOQPLSMILHF 942
DB 930 kysfmaaylclayrtlsgtetllshpeltclafnarhgvvvrvsgmays---llsnl 985
QY 943 QYHG 946
DB 986 evyg 989

```

```

RESULT 20
W88422
ID W88422 standard: Protein: 918 AA.
XX

```

```

AC W88422:
XX 26-APR-1999 (first entry)
DT Chlamydia pneumoniae surface exposed protein Omp9.
XX Chlamydia pneumoniae surface exposed protein Omp9.
DE Omp9: outer membrane protein 9; surface exposed protein; antigen;
XX Infection; diagnosis; vaccine; atherosclerosis; asthma.
XX Chlamydia pneumoniae.
XX W09858953-A2.
XX 30-DEC-1998.
XX 19-JUN-1998; 98WO-DK00266.
XX 23-JUN-1997; 97DK-0000744.
XX (BIRK/) BIRKELUND S.
XX (CHIRI/) CHRISTIANSEN G.
XX Birkelund S, Christiansen G, Knudsen K, Madsen A;
XX Mydind P;
XX WPI: 1999-105610/09.
XX N-PSDB: X06821.
XX
XX Claim 7: Page 56-58; 115pp: English.
XX
XX This polypeptide comprises the novel 96.7 kDa surface exposed
XX protein Omp9 of the human respiratory pathogen Chlamydia
XX pneumoniae. Its amino acid sequence was deduced from DNA (see
XX X06821) isolated from a C. pneumoniae expression library. The
XX invention provides 12 novel surface exposed proteins, Omp4-Omp15
XX (see W88417-28), and nucleic acid sequences encoding them (see
XX X06816-27). A new species specific test is claimed that is used
XX to identify mammals (including humans) infected with Chlamydia
XX pneumoniae. The test comprises detecting antibodies specific for
XX Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
XX membrane proteins, especially by PCR. The proteins are also used
XX in the diagnosis of C. pneumoniae infection in mammals. The
XX nucleic acids and proteins can also be used in the immunization of
XX mammals, the nucleic acids being particularly useful as DNA
XX vaccines for effecting in vivo expression of antigens. The
XX vaccines may also prevent atherosclerosis and bronchial asthma,
XX which are possibly associated with C. pneumoniae.
XX
XX Sequence 918 AA:

```

```

Query Match 12.5%: Score 637.5; DB 20: Length 918;
Best Local Similarity 27.2%: Pred. No. 8.7e-39;
Matches 262; Conservative 132; Mismatches 319; Indels 251; Gaps 47;

```

```

QY 61 LRYILAILOKTPNEGAAVTIDVLSFED---TQKRGIVFAKULTPESGGAICVYASPNP 116
DB 130 lrmhaa--ptlqk9-alkitdglvlesignldne-----nasenggal-----ntk 175
QY 117 TVEIRDTIGVIFENNTCCPFTSSNPNAVNNKIREGAIHAQ-NDYINHHHDVGVCFMKN 175
DB 176 tlslgtstrfvalgn-----sssgggaiaagdvlsenagjlsfgm 220
QY 176 FSTVRCGAISTANTFVYSENGSCFLEMNDICITNTNAGKGAIIYAGTSNFESENCDLFF 235
DB 221 satlsggalsaaenqlvismqn--liffgckatln-----ggal----- 257
QY 236 INNACGGAIFSPICSLTGNRCNIVFYNNRCFKNVETASEASDCGAIKVTTRLDVTCN 295

```


[illegible]

RESULT 15
B13639
ID B13639 standard; Protein; 1006 AA

QY	39	SLSKKISLITGTHLTCYLDNLKYLALAILOKPNKECAATITVDLSFDTOKKGGYFAK	98
Dh	78	nljsgfvljgphalt---fentl-----lstngaals-----nsaagf---	115
QY	99	NLTPESGCAICVSPNSPVEIRDTGPVLFENNTCCRPPTSSNP--NAAV-----	147
Dh	116	--ltiegfelstfscnslavl-----paatlckxgqrlpttstpngtlyaktldllln	168
QY	148	-----NKIR--GCAIHAONLEYININHHVYGKMKFYSVRGCAISTANTFVSENO	198
Dh	169	.nekfstysnlvsqgdgldakslvtqglsklvfgentaqdgagcqvltfsamaneap	228
QY	199	FLFMDNICIOTNTG--KGCAIYA-----GTSNSFE-----SNNCDLFIINNACCAG	243
Dh	229	laifanv-----agvrvqglaavqdgqgvsststledpvafstnlavelfgnarvg	282
QY	244	GAIFSPICSLTNGNINFIYNN--RCERN-----VETASSEASGCAI	284
Dh	283	gglvs-----ygnvaflnqgklflmvaspvylaakqpsqgsqntsmnygggail	334
QY	285	KVTRRLDVTGN-----RGAFESDNTIKNGCAIYAPVTVLDNGTYYINNANK	335
Dh	335	fcknagqagsgnsgsvsldqevglvftsnvaagkygailakkslvancqpvqcllrlind	394
QY	336	KCGAIYIDGTSNKRISADRAHAIIFENENI-----VYNVYNACTSTANPPRRNITVA	391
Dh	395	-ggailysgejelstlsadgdlldglnlkrakenaadvytlvs-----qdlsmgsg	448
QY	392	GEILIGACSSONLIFYDPIEVSN-----AGVSSTNKEADOTGSVSEGATVNSDFH	444
Dh	449	kltlrakaghlflndprieamanqnpagqskllkindgeyglcdvifan---gssltly	505
QY	445	ORNLQTKTPARLTNSNGFLCIEDHAIOTVNFOTGCVSNGCAVLSCKKNAGANS--A	502
Dh	506	qn-----vltiegqlvirekakisvnslsqglslymeagstldfvtpprpqpra	556
QY	503	SNASITTLKHIGLUNSSJLSCAGELPRLWEPRTN--NSNNYTAOTATFSLDVKLS---L	557
Dh	557	anqlitlslnlhslslallana---vlnpnlnpqadshpavlgst--tagvslspilf	611
QY	558	IDDYGNSPEYSTD-----LTHALSSOPMLISFASPDNRSDMDPFGSLNPNHG	607
Dh	612	fedloddctdydwlgisngklnvklqgltkpr--pnapssdltl-----gnempkyg	661
QY	608	WQGLTWGMAKTODEPPASSATTIDPOKANFPIHRTLLTLWLPAGVSPKHKRSPFLANTL	667
Dh	662	yggswklaw-----drranagpylkalwtelklynpqprelvaslvpnsl	706
QY	666	WGNMMLLATESLKSNAELPRSDHPFW--GITGGIGIMVYUODREHNPCHMSSGYSAGMI	726
Dh	707	wss--lldlrtsahsaigavsdgrtsrclglwvaganffhdaldaggyrylsgyslsg--	766
QY	727	AGQHTFSLKFSQYTKLNERAYAKNNV--SSKNVSCOGEMLFSLQEGFLTKVLGYSYG	784
Dh	764	-ansyfgsasmfjaeftefvg--skdyvvucfnhnacltgyvslstqca---lqgsylfg	816
QY	785	DINCHNHFTYOG--ENLTSGCTFRSOT-----MCGANFQFLPKPKPGSTHILT--APFL	833
Dh	817	dfcifaasy9fgqgmktvstvlseesdvwadmncclagealgaalprlvilpskylhnelprlv	878
QY	834	GALGIVSSLSHPTVECAVPRSFSTKTPRLINLVLPICVKGSGFMNATOPQAAVTELAYOPV	893
Dh	877	gqefsyadhesiteeqdqrakfs--ghllnlsprvykldfcisst--hpnkyslmaaylcd	934
QY	894	LVRQEGCITOLLAKSIGTIFGSGSPSSHHANSYKISQOTOPRLSMYTLHFQYNG	946
Dh	935	ayrltvisgretcllnqetwctdaflntrgvuvvgmsmas-----ltnieyvg	983

RESULT	16
Y16737	
ID	Y16737 standard; Protein; 1013 AA
XX	

```
Db      455 gklttlrekagbqllfndplemangnqpasepikindgeytgdlvf--fanganstly 512
OY      444 HORNLQTKTPAPLILSLNGFLICIEDHAQUTVNRFTOTGGVSLGCAVLSCKKNAGNS-- 501
Db      513 qn-----vlieqgrlvrkaklsvnsiqsgslymeagstldvrlcpqpqp 562
OY      502 ASNASITLKHIGLNLSSILKSGAEPLLMWPTN--NSNNYTAADTAATFSLSDVKLS--- 556
Db      563 aaeqlitlnhlslslslanna-----vtnptnppagdsnpavlgst-tagprrtisgpf 617
OY      557 LIDYGNSPYESTD-----LTHALSSQPMLSISEASDNLRSDDMFSGULNPHY 606
Db      618 ffeiddtdlaydrydgwngskldvklqqlqfqp--saapadlrl-----gnempky 667
OY      607 GMOGLMTMGMAKKTQDPERASSAATITDPOKANRPHRTLLTWLPAGYVSPKHRSPLIANT 666
Db      668 gygswklaw-----dptlanapylkalwklcqpnpqpervasiypns 712
OY      667 LMGNNLLATESLAKNSAELTPSDHPEW-GITGGGLGMVYQDPRRNHPGFIHRSSQYSAGM 725
Db      713 lwgs-ildirahaaigavsgvgrglwsgvsnfgyhdardalgqgyrlsggyslg- 770
OY      726 TAGOTHTFSLKFSQTYTKLNERVAKNNV--SSKNYSCGEMLFSLQEGFLLTKLVGLTSY 783
Db      771 --ansyfgssmfglaflevfgr-ekdyvvrsernhacisgsvylstkqg-----lcsyylf 822
OY      784 GDHNCNHHRYTQG-ENLVSQGTFRSQT-----MGAVFPDLPKPKPGSHILIT--APF 832
Db      823 gdafirasvgfngqumktsyltaeesdvrdwncnclvgeilvgvrlpvtprsklylnelpf 882
OY      833 LGALGATSYSLHFTFVGAVPRSFSTKPLINVLVYIGVKGSEFMATQPOAMVELAYOP 892
Db      883 vgaeisysdhesfteeqgdafatfrs-qhlmlslvpygkfdrcsst-hpnkysfmgaylc 940
OY      893 VLYROEPIATQLASKGIWFGSGSPSRHAKSYKISQOTOPLSLTLHFQYHG 946
Db      941 dayrtisgltqllshgqelwtldaftharhylvirgsmas----ltanleyg 990

RESULT 13
Y69369 standard; Protein; 918 AA.
XX      AC      Y69369.
XX      DT      19-JUN-2000 (first entry)
XX      DE      Amino acid sequence of the CPN100395 polypeptide.
XX      KW      CPN100395; Chlamydia infection; immune response; vaccine.
XX      OS      Chlamydia pneumoniae.
XX      PN      W0200011183-A2.
XX      PD      02-MAR-2000.
XX      PE      18-AUG-1999; 99WO-1B01449.
XX      PR      20-AUG-1998; 98US-0097187.
XX      PR      20-AUG-1998; 98US-0097188.
XX      PR      20-AUG-1998; 98US-0097189.
XX      PR      20-AUG-1998; 98US-0097190.
XX      PR      20-AUG-1998; 98US-0097195.
XX      PR      20-AUG-1998; 98US-0097196.
XX      PR      20-AUG-1998; 98US-0097197.
XX      PR      27-AUG-1998; 98US-0097191.
XX      PR      17-AUG-1999; 99US-0376770.
XX      PA      (CONN-) CONNAUGHT LAB LTD.
XX      PI      Mordin AD, Oomen RP.
```

```
XX      DR      WPI: 2000-224703/19.
XX      DR      N-PDB: 261509.
XX      PT      Novel antigens and corresponding DNA molecules that can be used to
XX      PT      prevent, treat and diagnose disease caused by Chlamydia infection in
XX      PS      mammals, especially humans -
XX      PS      Claim 19; fig 15-E; 201pp; English.
XX      CC      Y69362-69 represent Chlamydia pneumoniae polypeptides. The polypeptides
XX      CC      are present in the bacterial membrane structure. The polypeptides
XX      CC      vicinity of the membrane structure, in the inclusion membrane structure,
XX      CC      structure, in the external vicinity of the inclusion membrane structure,
XX      CC      and in the cytoplasm of the infected cell. The polypeptides may be
XX      CC      used to prevent, treat and detect the presence of Chlamydia infection
XX      CC      and/or the presence of Chlamydia in a sample. The polypeptides may
XX      CC      also be used to induce an immune response in a mammal. The vaccine
XX      CC      vector comprising the polynucleotides is used to induce an immune
XX      CC      response in a mammal. Antibodies directed against the polypeptides
XX      CC      may also be used therapeutically to treat and/or prevent a Chlamydia
XX      CC      infection.
XX      SO      Sequence 918 AA:

Query Match          12.8%; Score 649.5; DB 21; Length 918;
Best Local Similarity 27.2%; Pred. No. 1,1e-39;
Matches 261; Conservative 134; Mismatches 322; Indels 243; Gaps 46;

OY      61 LRYTALLQKTPNCGAAVTITDYLSFPTQKEGIFAKNLTPESGALIGVSPSPYEI 120
Db      130 lrmiaa--prtlgy-alkitdglvf--esignldlneassengal-----tktlsl 179
OY      121 RDTIGPVTFENNTCCRPFTSSNPNAAVNKIREGAIHAO-ULYINHNHDVGFPMKFSYV 179
Db      180 lgrstfvallgn-----ssqggallvasgdsvalsenaglsifmnsatl 224
OY      180 RGAISTANTFVSENDSCFLFMDNICIOTNPAKGAIVAGTNSFESNCDLFFINNA 239
Db      225 sgaisaegnlvisnqn--lffdgckatln-----ggaal-----d 258
OY      240 CCAGATFSPTCSLTGNAGNIVFYNNRCFKNVERTASSASOGCAIKVTRLDVGNRGRI 299
Db      259 ckgagapdpdlitlsgne-sihflnn-----tagngsgal-ykklvlvssqrgv 306
OY      300 FFSQNTKNNCGAIVAPVTLVDNCPYFINNIANNKGAIVIDTSSKRSKISADHAIIF 359
Db      307 lfsnkaan-----alpkgaialldsgelsisadlgnlff 342
OY      360 NENIVTNANGSTSANPP--RRNAITVASSGCEILLGAGSSONLIFYDPIEVSNAKV 417
Db      343 -----egntstlsgspasvtrnaidlasmakflnlratrgkvlyfypitssgatl 393
OY      418 SVSEFKKAD-----QTGSVVFSGATVNSADPHQ-RNLOTKTPALITSLNGFLICIEDHAQ 470
Db      394 klslnk-adagsgnltyegylvfygsklseeejlkpndnlkstftgaveiaaqaylvkdqvt 452
OY      471 LTVNRFOTGCVVSLGNGAVLSQYKNGAGNSASNSITLKHIGLNLSSILSGAEPLLM 530
Db      453 vvanliltlyegskvymdg-----gltfesaeagvltlngalidsl----- 493
OY      531 VEPITNNSNNYTAADTAATFSLSDVKLS---LIDYGNSPYESTDITLHALSSO---PMLSI 583
Db      494 -dgtnk-----alikataaskdvalsgpmlvdagyny-yehe---hmlsqgqvlyllei 542
OY      584 SEASDNLRSDDMDFSGULNVP-HYGMQGLMTMGMAKKTQDPERASSAATITDPOKANRPHRT 642
Db      543 s-agqtmctldipdplintlnhygyqgnwlvw-----datekt-----kn 585
OY      643 LTLTWLPAGYVSPKHRSPLIANTLMGN-----MLLATESLAKNSAELTPSDHPEW 692
Db      586 altltwtkykpnpetqgplvpnsllwsgstvdvrsiqslmdrstslssstlnlws----- 640
```


Mon May 7 09:06:31 2001

us-09-677-752-2.rag

Page 8

[illegible]

Db	176	ldsytlmnlgeqgflidhvsknrggalvygvsldltdhlpvlkknqleedssf--gggl	233
Oy	219	yactsnsefnennccdlffinnaccagalfstycslgtcnrgniivyrnnrcfnenvetassea	278
Db	234	fcraevn--lerynqlqindnasggvvyllpdyvlisnkelieishsaaasintlaagkl	292
Oy	279	sdgailvtttrldvtgn--rgriiffssdnitknvgaiy----apvvtlvdngprryfinni	332
Db	293	ypgggimctslshempkglif--nnktaelsggvylcrdlsssktlv-----rtafims	346
Oy	333	annngcaiyidg----tsnksiadrrhaiifennivvtvnanotstsnpprrnativa	388
Db	347	atsggalinsglsyctqgnfflaeadydlilnmhti-----tsspggyrnal_ya	397
Oy	389	ssscetillgacssncllfydprievsnacvs--vsfnkeadotgsyvfscatvnsadfhorn	447
Db	398	apgnlkrigargykillfydpdldnqtdldvlfnyephnhlgtvlfsqindsnatnpln	457
Oy	448	lotrtpavltsnqfleciedhaolvtvnffotgcvsvslngcavslscyxngcngnsasnsai	507
Db	458	flskfmsrrlervlaleodaaascctlsqvggllrlgnaallf--tkpgp-----ssi	510
Oy	508	tlrkilgnlssllksgaeirlplwvptnnsnnytdratatslsdvklslddydgsnye	567
Db	511	nfnalaahlpblseaseapkfwiylptlgtstysedstsltlsg--pltlindennpyd	569
Oy	568	stdlthalssopmlsisasdnqlrsdmdrsglanv-----phynoglwttwgaakrod	621
Db	570	sldl-----seprldlpprlpprcdckhltsnllveeamldehygygqfwsywmel--	622
Oy	622	pepassatittpokanrpfhrtlltlwlpagvyppskhrspilantlm-----gnmlt	673
Db	623	---cttscstvpqgncnhrqlgyvdwcpvygrrpmpertigetantlwgssaynallqitll	679
Oy	674	ateslkn---saeltpsdhpfwgitgglgkmvnyodprenhpgfmrssgysagmia--g	728
Db	680	ppqnlkxendleasld-----glllindnregckgfrnhttyeaattsakta	727
Oy	729	otwrtslkrfscotyrklmrkryaknnvssknyssocqemlfsldbgfltlkylglsygdhn-	787
Db	728	arhsflslqiaqmiskrfetrgspstltsnhylagllfdslllfrdlstlslglsygdhnm	787
Oy	788	-chhrrytogentlscgtfrrsotmggavffdlpkrpfgsthl--tapfgalgiysslshf	845
Db	788	lch--ytelkygsskaefnmhtlvasl--declfparrltrelcplgfisajalrsgasf	843
Oy	846	tevgaypsfsftrkplnlvlpigvksfmatnorpomavtelayovulyroepciatql	905
Db	844	qetghlirkfhpklpdlldlsspifgrsewkshhpmjwlteislvypltyknpmfll	903
Oy	906	laskaiwsgsgspssrrhamstyikisootoprlsmultlhyohgryssstpcnylncge	960
Db	904	linsglwtlqtqhapvynsvaaklnhtnslqllfsrvtlslsdysaqvssstlvgyklae	958
RESULT 9			
id	y92830	standard: Protein: 954 AA.	
xx	y92830:		
xx	29-AUG-2000	(first entry)	
xx	C: pneumoniae CPN100662 antigen.		
xx	Antigen: anti-inflammatory: respiratory: antibacterial: anti-asthmatic:		
xx	anti-arteriosclerotic: vaccine.		
xx	Chlamydia pneumoniae.		
xx			
xx	WO200024765-A2.		

```
Db 616 klllyanwsplygphpergetflnalwgsayatalaglhiss-----wdeekyha 667
Oy 698 ----GLCMWYODPRENHPCGHHMRSSGYSAGMTA--GOTHTFSLKFSQOTTKLNERAKN 751
Db 668 aslqglilvqkdkngfkfgrshmlqysaltcaatsqpsrlqfqlskakehesqn 727
Oy 752 NVSSKNV---SCGEMLFSLQEGFLTKLVGL-YSGDHNCNHFYTOGENTISQGTFRSQ 807
Db 728 stshhyfsgmciemtlf--kewirls--vsajymfesehhtmyqgllegnsgsfmnh 783
Oy 808 TMGCAVFFDLPMKPFEGSTHILTAFLGALGYSSLSHTEGAVPRSFSTKPLINLVLP 867
Db 784 llaqalscvflpqpghes-lqlypfitalaigrnlaaqesgdharefslhrpltdvalp 842
Oy 868 IGVGSEPMATQROQAWTVELAYOPVLARQEPGATATQLLASKGIWFGSGSSRHAMSYK 927
Db 843 vglraswkhhrvplvwlteisyrsflyrqdbelshskllsqglwtlqatpvtynalqik 902
Oy 928 ISQOTPLSMTLHFOYHGFYSSSTFCNYLNGEIALRF 965
Db 903 vknemqvfpkvtlslsdyssdlssstslshynvasrmrf 940

RESULT 6
Y35083
ID Y35083 standard; Protein: 969 AA.
AC Y35083;
XX
DT 13-SEP-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed polypeptide.
XX
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.
XX
OS Chlamydia pneumoniae.
XX
PN W09927105-A2.
XX
PD 03-JUN-1999.
XX
PE 20-NOV-1998; 98WO-1B01890.
XX
PR 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97ER-0014673.
XX
PA (GEST ) GENSET.
XX
PI Griffiths R;
XX
PT WPI: 1999-357842/30.
XX
PS Genome sequence of Chlamydia pneumoniae
XX
Page 967-969; Disclosure: 1912pp; English.
XX
Y3584-Y35879 represent the proteins encoded by all the open reading
XX frames in the complete genome (see X91990) of Chlamydia pneumoniae.
XX C. pneumoniae causes respiratory disease such as pneumonia and
XX bronchitis and is thought to be a contributing factor in heart
XX disease, sarcoidosis, sinusitis, purulent otitis media, erythema
XX nodosum or pharyngitis. The polypeptides encoded by the open reading
XX frames of the C. pneumoniae genome (see Y34584-Y35879) can be used in
XX immunogenic compositions as vaccines. Vectors containing C. pneumoniae
XX nucleotide sequences can also be used as immunogenic compositions,
XX especially where the vector directs the expression of a neutralising
XX epitope of C. pneumoniae.
XX
SQ Sequence 969 AA:
```

```
Query Match 20.8%; Score 1057.5; DB 20: Length 969;
Best Local Similarity 30.5%; Pred. No. 5.2e-70;
Matches 302; Conservative 165; Mismatches 447; Indels 77; Gaps 28;

Oy 6 FFELIGNSLGLAREVPSRIFLMNSVDP-----PKRESLSNK-----ISL 46
Db 25 fllylgnfltaemigtcpavyslqdslekfalerdeefrsfpildslstlgtfspiltf 84
Oy 47 TGDTHNLT-NCYLDNLRLVILALLOKTPNEGAAVITQVLSFDTQKGYIPAKULPESG 105
Db 85 vgnrhnsqgdvlshysxldnlllwtsaagavscnml--lshvedhafsklalaigt 142
Oy 106 GAIGVASPNSPVEIRDTIGPFIENNTCCRPFTSSNPAVANKIREGAIHAQ-NLYIN 164
Db 143 galacqg----actlknrgpliffnrgln-----nastlgetrgegalacngdtlis 191
Oy 165 HHNDVGMKKNFSTYVRGCAISTANTFVYSENOCFLEMDNICIOTNAGKGAITYACTSN 224
Db 192 qnqglfyfvnsvnmwggalslmgchcrlqsnrabilfn-----nlapsqgalaissent 245
Oy 225 SPESNNDLFEFINNACGAGALFSPICSLTGNRGNIVFYNNRCFKNVTASSEASDGA 283
Db 246 ltsdnlrplylknmcgmnggalqtsvalknsgsvlfnnlalsg--sinsngsgga 303
Oy 284 IKVTRLDVTGNRGRIFESDNTITKNYGAIYAPVTVLDNGPTFYINNANKGAIYID 343
Db 304 l-ytenlsiddnpgclifnmnyclrdgagictqglfiknsghyvftnn-qgnwggalml 361
Oy 344 GTSNSKISADRHAIIFNENIYTNWTNANGTSTSANPPRRNAITYASSGELLGAGSSQN 403
Db 362 qdstclllaegnalafqnevflltfg-----fynaahclups-nlqlganqyt 410
Oy 404 LIFYDPIEVSNAGVS-VFENKEADQTSVVFSGATVNSADRHORLQKTPAPLTLNCGF 462
Db 411 tafidpqlhpbctnplifnpanngcliflssaylpaasdyemfisskntseelrgv 470
Oy 463 LCIEDHAQLTVNRFOTGTGGVYSLNGAVLSCYKNGAGNSAS-NSITLKHIGLNLSTLK 521
Db 471 lstedragwqfykftqkgjllkghaaslattnseptsavsgvllnnlalnplsia 530
Oy 522 SGAETPLMVEPTNNSNNYTADTATFSLSDVKLSLIDYGNSPRESLDLHALSSQML 581
Db 531 kg-kapllwlrplqgsapfltednpltltsq-rltlnenrdydsldseplnlll 588
Oy 582 SISEASDNLNSDMDPFGLN-VPHYGQGLMTWGMATQDPEPASATITDPOKANFH 640
Db 589 slsdvlarhlnclnlpeslnatehygyglwspwyvet--littnasi---etancly 643
Oy 641 RTLLTWLRAGYVPSPKHRSPLIANTLWGNMLLATESLKSNAELTPSD--HPFWGTGGG 698
Db 644 ralyanwtrplykknvpeygdlatrlwgsfhtmtsllrsytrtdgsdlerplelqgia 703
Oy 699 LGMWYODPRENHPCGHHMRSSGYS--AGMIAGOTTFSLKSQOTTKLNERAKNNSK 756
Db 704 dglfwhqnsipgarpfirlstlqyslqasasetshkklslgtqflltrkclkyssnmvsaah 763
Oy 757 N--YSCGEMLFSLQEGFLTKLVGLYSGDHNCNHFYTOGENTISQGTFRSQGTGAVF 814
Db 764 ntvsalylwelp--fgeaf-atsvtlayygdhlnslhpsnqe-qaegtvyshltaaal 820
Oy 815 FDLPMKPFEGSTHILTAFLGALGYSSLSHTEGAVPRSFSTKPLINLVLPICVKSF 874
Db 821 cefpqqksyhl--spfgatairshqtafeelknfkvsgpfytlrlplrlqgkw 878
Oy 875 MATQROQAWTVELAYOPVLARQEPGATATQLLASGIGWFGSGSSSRHAMSKISQGTQ 934
Db 879 gskflvprlewletlsyepilyqnpqilyvllaasgswdlghnyvrnalgykknvntal 938
Oy 935 LMSWTLHFOYHGFYSSSTFCNYLNGEIALRF 965
Db 939 frsldflidygsvsstethlqagstlkf 969
```


XX Disclousure: Page 999-1000; 1755pp: English.
PS Y36754-Y37949 are encoded by open reading frames (ORFs) of the genome of
CC Chlamydia trachomatis (see 201425). The polypeptides can be used as
CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC can also be used to control growth of the microorganism. Chlamydia
CC trachomatis is responsible for a large number of diseases, e.g. eye
CC diseases such as conventional trachoma, nongonemic trachoma,
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
CC perihepatitis, Bartholinitis; pneumonia; in breast feeding infants;
CC and venereal lymphogranulomatosis. The polypeptides of the invention
CC may be of use in treating these diseases.
XX
SO Sequence 989 AA:

Query Match 99.2%; Score 5047; DB 20; Length 989;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 958: Conservative 4; Mismatches 3; Indels 0; Gaps 0;
OY 1 MKKAFFFLLNSLSGLAREVPSRFLFMPNSVDPPTKESLSNKISLTGDTNNTNCLYLDN 60
DB 25 MKKaffllllsgslsrlarevpsrflfmpnsvdpptkeslsnksltgdtlnlnclyldn 84
OY 61 LRYTILAIQKTPNEGAAVTITDYLSPFDTOKEGIYFAKNLTPREGGATGYASPSPTVEI 120
DB 85 lrytllalqktpnegaaavttitdylsfldtqkegilyfaalnlpesggalyaspsptvei 144
OY 121 RDTIGPIYFENNTCCRPFTSSNPAAVKKIREGAIHNAONLYINIMHNVGFMKNEFVVR 180
DB 145 rdtigpylfemntccrpfstspnaavknkiregailnaqnlyinhndvvgfmknfsvvr 204
OY 181 GGATSTANTPVSNOSCFIFMDNTCIGTNTAGGAIYACTSNPSFSSNCDLFFINNAC 240
DB 205 ggastantlvsnoscfifmdntcigtntaggaiyactsnpsfssncdlffinnac 264
OY 241 CAGGATFSPICSLTGNRGNIVFYNNRCFKNETASSEADGAIKVTTRLDVNTNRGRIF 300
DB 265 caggatfspicsltgnrgnivfynnrcfknetasessdggaikvtrldvntnrgrif 324
OY 301 FSDNITNNYGAIAPVYTLVDNCPYFINNANNGGAIYIDGTSNSKISADRHAIIFN 360
DB 325 fsdnitknysgaiajapvylvdnqpyffinnannkgaiyidgtsnksisadrhailfn 384
OY 361 ENITVNTNNGSTSNANPPRRNAITVASSSGELLGAGSSQNLIFPDPIFVSNAGVSVS 420
DB 385 enitvntnngstsnanpprrnaitvasssgelllgagssqnllfydpilevsnagvsvs 444
OY 421 FNEADDTGCVNFGATVNSADFHORNLQTKTPAPLTLNSGFLCIEDHAQLTVNRFTQNG 480
DB 445 fneaddtgsvnfgatvnsadfhgrnlqtktpapltlntsgflciedhaqltvnrfqtg 504
OY 481 GYVSLGNGAVLSTCYKNGAGSASNASITLKHLNLSTLKSQAELPDLWVEPTNNSNNY 540
DB 505 gyvslngnavlscyknagagsasnasitlkhiglnlsslksgaepldlwveptnnsny 564
OY 541 TATTAATFSLSDVKLSLIDYGNSPYESTDLTTHALSSQPHLSISEASDNOLKSDMDPSG 600
DB 565 tattaatfslsdvklslidygnspyestdlthalsqphlssiseasdnolksdmdpsg 624
OY 601 LNVPHYGMQGLMTWGAKTODPEPASATITDPQKANRFRTLLTLTWLPAGVYSPKXHS 660
DB 625 lnvphygwqglwvgaaktodpepasatitdpkkanrfrtlltlwlpagvyspkxhs 684
OY 661 PLANTLMLGMLLATSLSKNSAELTSPDHPFWCITGGGLGMVYDOPRENIHPCFHRSSG 720
DB 685 plantlmlgmllateslksnsaeltpsdhpfwctggglgmvydoprenihpcfhrssg 744
OY 721 YSAGMIAGOHFHSLSFOSQYTKLNERVAKNNVSSKNSCOGEMFLSLOGEFLTLTKVLG 780
DB 745 ysaqmiaagqthfhsifsqlytklnervayaknnvssknscoqemflslogellakvl 804

OY 761 YSYGHNCHHEFTQGENLTSGCTFRSQTMGAVFEDLPKPGSTHILTAFLGALGIS 840
DB 805 ysygdnchhfytqgentlsgctfrsqtmgavfddlpmpkpgsthilltaplgaiglys 864
OY 841 SLSHTEVGAVPRSRSTTPLNLVPLPGVGSFNNATROPAWTEALAYOPVLYROEPG 900
DB 865 slshtevgaiprststplnlvplpgvgsfnnatropawtealayopvlyrgele 924
OY 901 IATQILASKGIMFGSGSPSSRRHMSKRSOOTOPLSMWTLPHOYGEFSSSTFCVYLNGE 960
DB 925 iatqlaaskgimfgsgspssrrhmskrsqqtpiswiltlphqnglyssstfcvylnge 984
OY 961 IALRP 965
DB 985 ialrf 989
RESULT 2
BI3634
ID BI3634 standard; Protein; 964 AA.
XX
AC BI3634;
XX
DT 02-FEB-2001 (first entry)
XX
DE C. trachomatis pmpe gene protein.
XX
KW Chlamydial infection; sexually transmitted disease;
KW pelvic inflammatory disease; PID; tubal obstruction; infertility;
KW trachoma; blindness; acute respiratory tract infection;
KW atherosclerosis; coronary heart disease; antibacterial.
XX
OS Chlamydia trachomatis.
XX
PN W0200034483-V2.
XX
PD 15-JUN-2000.
XX
PF 08-DEC-1999; 99MO-US29012.
XX
PR 08-DEC-1998; 98US-0208277.
PR 08-APR-1999; 99US-0288594.
PR 01-OCT-1999; 99US-0410568.
PR 22-OCT-1999; 99US-0426571.
XX
PA (CORI-) CORIXA CORP.
XX
PI Probst P, Bhatia A, Skeiky YAM, Fling SP, Jen S, Stromberg EJ;
XX
DR WPI; 2000-431303/37.
XX
PT Isolated polypeptide for diagnosis and treatment of Chlamydia infection
PT comprises immunogenic portion of Chlamydia antigen, which comprises
PT amino acid sequence encoded by polynucleotide sequence
XX
PS Claim 2; Pages 184-186; 256pp; English.
XX
CC The present invention relates to new nucleic acid sequences and the
CC proteins encoded by the nucleic acid sequences. The encoded proteins
CC comprise an immunogenic portion of a Chlamydia antigen. The encoded
CC proteins are useful for the serodiagnosis and treatment of Chlamydia
CC infection. Chlamydiae are intracellular bacterial pathogens that are
CC responsible for a wide variety of human infections. C. trachomatis
CC infection is one of the most common sexually transmitted diseases and
CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction
CC and infertility. Trachoma due to ocular infection with C. trachomatis is
CC the leading cause of preventable blindness worldwide. C. pneumoniae is a
CC major cause of acute respiratory tract infections in humans and is also
CC thought to play a role in the pathogenesis of atherosclerosis and
CC coronary heart disease. The present sequence is a protein isolated in the
CC present invention.
XX

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: May 6, 2001, 19:22:01 ; Search time 22.61 Seconds

(without alignments)
2668.675 Million cell updates/sec

Title: US-09-677-752-4

Perfect score: 4533

Sequence: 1 MRPDHNNFCCICAAILSTPA.....CVLRGSGSHSYLDLGTTRF 878

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4509	99.5	878	2	B71460
2	3543.5	78.2	867	2	F81721
3	947.5	20.9	928	2	B72077
4	919.5	20.3	930	2	A81591
5	915.5	20.2	930	2	D72078
6	894.5	19.7	1276	2	C81591
7	894	19.7	1407	2	B72078
8	883	19.5	928	2	G81591
9	845	18.6	928	2	D72077
10	845	18.6	949	2	F81591
11	816.5	18.0	973	2	F72076
12	816.5	18.0	995	2	C81593
13	773	17.1	936	2	C72078
14	772	17.0	936	2	B81591
15	758	16.7	922	2	H81722
16	758	16.7	1013	2	G71460
17	740	16.3	922	2	B72131
18	739	16.3	922	2	F81539
19	708.5	15.6	841	2	E72130
20	575.5	12.7	946	2	D81594
21	574.5	12.7	946	2	C72075
22	566	12.5	947	2	D72067
23	557.5	12.3	978	2	B81593
24	556.5	12.3	978	2	G72076
25	540.5	11.9	938	2	H72074
26	531	11.7	934	2	A72075
27	531	11.7	952	2	D81593
28	510.5	11.3	1520	2	A81731
29	503.5	11.1	1609	2	H72013

30	493.5	10.9	986	2	B81675	polymorphic membra
31	487	10.7	1016	2	H71460	probable outer mem
32	484.5	10.7	975	2	F71518	hypothetical prote
33	482.5	10.6	976	2	F81722	polymorphic membra
34	475	10.5	964	2	E71460	probable outer mem
35	463	10.2	983	2	A81723	polymorphic membra
36	453	10.0	1531	2	H71460	probable outer mem
37	451	9.9	1034	2	F71460	probable outer mem
38	449.5	9.9	1025	2	G81722	polymorphic membra
39	445.5	9.8	1732	2	C81601	polymorphic membra
40	444.5	9.8	1723	2	E72067	polymorphic membra
41	413	9.1	1672	2	C81675	polymorphic membra
42	391	8.6	1751	2	G71518	hypothetical prote
43	341	7.5	1460	2	B81675	polymorphic membra
44	339.5	7.5	514	2	E72076	polymorphic membra
45	287.5	6.3	1770	2	A71517	hypothetical prote

ALIGNMENTS

RESULT 1
B71460
Probable outer membrane protein I - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: B71460
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A>Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A:Reference number: A71570; MUID:99000809
A:Accession: B71460
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-878 <ARN>
A:Cross-references: GB:AE001361; GB:AE001273; NID:g3329348; PIDN:AAC68472.1; PID:g332
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: pmp1

Query Match 99.5%; Score 4509; DB 2; Length 878;
Best Local Similarity 99.5%; Pred. No. 5,76-277;
Matches 874; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY	1	MRPDHNNFCCICAAILSTPAVLFQGDPLGETALLTKRPNHVVCTFFEDCTMESLPALCA	60
DB	1	MRPDHNNFCCICAAILSTPAVLFQGDPLGETALLTKRPNHVVCTFFEDCTMESLPALCA	60
QY	61	HASODPLVVLGNSYCWPFVSKLHTDPKELFKKGGDLSIONFRFLSTDCSSKESP51	120
DB	61	HASODPLVVLGNSYCWPFVSKLHTDPKELFKKGGDLSIONFRFLSTDCSSKESP51	120
QY	121	IQHNGOLSTIRNNGSWSEFCRNHAECSGAIISADAFSLQHNVLFTAFEESSKNGAIOA	180
DB	121	IQHNGOLSTIRNNGSWSEFCRNHAECSGAIISADAFSLQHNVLFTAFEESSKNGAIOA	180
QY	181	QTFSSLRNVSPISEFARNRDLNGAICCSNLICSGVNNLFTFGNSATNGAICCSIDLN	240
DB	181	QTFSSLRNVSPISEFARNRDLNGAICCSNLICSGVNNLFTFGNSATNGAICCSIDLN	240
QY	241	TSEKGSLSLACNOETLFFASNAKKEGAIYAKHMYLRYGAPVSTFINSKIGATAIOG	300
DB	241	TSEKGSLSLACNOETLFFASNAKKEGAIYAKHMYLRYGAPVSTFINSKIGATAIOG	300
QY	301	GSLSLIAGGSVLFQNNQSORTSDOGLVRNAYILEKDALISLEARNGDLFFDPIVOESS	360
DB	301	GSLSLIAGGSVLFQNNQSORTSDOGLVRNAYILEKDALISLEARNGDLFFDPIVOESS	360
QY	361	SKESPLSSLSQSVSPPTATASPLYIOTSNARSVIFSSERLSEEEKTDNLTSQLQPT	420
DB	361	SKESPLSSLSQSVSPPTATASPLYIOTSNARSVIFSSERLSEEEKTDNLTSQLQPT	420

Query Match 20.9% Score 947.5; DB 2: Length 928;
Best Local Similarity 31.6% Pred. No. 7.2e-52;
Matches 288; Conservative 137; Mismatches 336; Indels 149; Gaps 32;

```
85 TDDEKALFKREKGLSIONFNF-LSPFDCCSKESSPSIIHOKNG-QLSLRN----- 132
Db 52 TNDGTTNLTGDSITNMACSPALTAACRKTETGNLSPFGHGQFLQYIDAGANCTFT 111
133 -----NGSMSE-----CRNHAEGSGGAIADAFSLQHNLYLTFEENSSKNGGAI 178
Db 112 NTAAANLLSSGFSYLSLIOTTMATGTGAIKRTGACSIOSNV-SCFGQGNFNDNGAL 170
179 QAOFTSLRNVPISFARNRADLNGAICCSNLICSGN-VNPLEFTGNSATN----- 229
Db 171 QGSSISLSLMPN-LTFAKNKATOKGGLYSTGGITTNNLTNSASFSENTAANNGAIYTE 229
230 -----GGAICCSIDLNTSEKGLSLACQETLFAFNSNAKEK 265
Db 230 ASSFISSNKAISFINNSVATSTAGATYCSS--TSAPKPVLLSDGELNFIGNTATIS 287
266 GGAIVAKHMYLRNGVPSFINNS-----AKIGGAIQSGSSTILAGESSVLFQNNNS-- 318
Db 288 GGAITYDNLVLLSSGPTLFFKNNSAIDTAAPLGAIAIADSGISLSLAGDITFEQNTV 347
319 -QRTSDQGLVRNAIYL-EKDAILSSLEARNGD-ILFEDPIVOESSKESPLPSSLOASVT 375
Db 348 KGASSSOTTNRNSINIGNTAKIYOLRASOGNTIYFDPIT-----TSITLALS 396
376 SPTPATASPLVIQTSANRSVIFSESERLSEERKT-PDNLTSOLOOPTIELKSGRLVLMKRAV 434
Db 397 DALNLGPPDLAPPAIQTGIVTFEGEKLSEAEALADNLKSTIOOPLTLAGOSLSKGV 456
435 LSNPSSLQDPPQALLMEAGTSLKTSDDLKATLSIPLSLDTEKSVYIHAPNLSIQKIFL 494
Db 457 LVAKSPISQSPGSLTMDAGTLETAAGTITNNLVANVDSLKEKTKATLKAQNAS-QIVYTL 515
495 SNS-----GDENFEYENVELLSKEQNN--IPLLTSLKEQ-SHLHLPDGNLSS----- 537
Db 516 SGSLSLVDPSEGNAYEDV-----SMNPNQVPSCLTLTADDPANIHITD--LAADPLEKNPI 568
538 HFTYQDWMTFESKMDSDGSHLIA--NMTPKNVYPHPEROSTIVANTLMNTYSDMQAOSM 595
Db 569 HMYQGNMWSLMQEDTATKSKAATLTWTGTGYNPNRKRGTLVANTLWGSFVDRSTIOQL 628
596 INTIAGGAVLFGTWSAVSNLFYANDSGKPIDNMHRSGLYLFISTSLDDHISFCLA 655
Db 629 VAKTVASQSETRGIMCEGINSNF--HKDSTRIKKGFRHISAGYVVGATTTLASNLITAA 686
656 AGOLLCKSSDSFTSTETTSYIATVOAQ-LAT-----SLMK-----ISAQACYN 698
Db 687 FCOLLEKDRDHFINKNKAASYAASLHLOHLATLSPSLRYLPGSESEQPVLEFAQISYI 746
699 ESTHELKTYSKRSKRGFGSMHVAASGEVCASTPIVS-NGSGLFSSFSIFSKIOGSGT 757
Db 747 YSNNTKMTYITQAPK-GESSWYNDGCALELASSLPHTALHHEGLFHAHYEPFIKEASYIH 805
758 QDSFEE-SSGEIRFSFASFRNISLPIGITFEKKSKQKTRFYVFLAYIODLKRQVSGP 816
Db 806 QDSFKENNTLVASFPSCGLINVSVPIGITFEKFSRNRASYETTYIVADVVRKKNDCJ 865
817 VLLKNAVSDAPMANIDSA-----YMFRLTNORALHRLQTLINVSQVLRGQSHSY 868
Db 866 TALLINNTSKRTIGTINLSRQAGIGRAGIIFYAFS-----PNLEVTSNLMEIRGSSRSY 918
OY 869 SIULGTTYR 878
Db 919 NADLGKGFQ 928
```

RESULT 4
AB1591

polymorphic membrane protein G family CP0307 (imported) - Chlamydia pneumoniae (S
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: AB1591
R:Read: T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
C:; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: AB1500; M0ID:20150255

A:Accession: AB1591
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1930 <REA>
A:Cross-references: GB:AE002193; GB:AE002161; MID:g7189234; PIDN:AAF38164.1; PID:g718
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0307

Query Match 20.3% Score 919.5; DB 2: Length 930;
Best Local Similarity 31.0% Pred. No. 4.3e-50;
Matches 260; Conservative 150; Mismatches 343; Indels 87; Gaps 26;

```
98 LSIQNFRLSF-----TDCSSKSSPSIIHOKNGQLSRNNGSMSPCR--NHAEGSGA 149
Db 119 LTFPGESNLSPFIAAGCTTVASGKSTLS-----SAGALVLTONGTILFSQNSVNEANNNGA 174
150 ISADAFSLQHNLYLTFEENSSKNGGAI-QAOTFSLSRVSPISFARNRADLNGAICC 208
Db 175 ITAKTSLISGNTSIFFTSSAKKLGAIYSSAAASISGNTQGLVFNNKGETGCGAL-- 232
209 SNLISGNNV--NPLEFTGNSATN-----GGAICCSIDLNTSEKGLSLACQETLFAFNS 261
Db 233 -GEFASISIQNSLSLFFSGNTATDAAGKGAICY--EKTGETPTLISGNKSLTFEANS 288
262 AKEKGAIVAKHMYLRNGVPSFINN-----SAKIGGAIQSGSSTILAGESSVLFQNF 316
Db 289 SVTQGAICAHGLDLSAGPTLFSNNRCNTAKGKGAIAIADSGISLSLAGDITFEIG 348
317 NS-QRTSDQGLVRNAIYLEKDAILSSLEARNG-DILEFDPIVOESSKESPLPSSLOASV 374
Db 349 NTLTSTAPSTRNAIYLGSSAKITNLRAAGOSIYYDPLASMTTASDVL----- 400
375 TSPPTATASPLVIQTSANRSVIFSEERLS-EERKTPDNLTSOLOOPTIELKSGRLVLMKRA 433
Db 401 TINOPDSNSPL--DYSGITVEFGEKLSADEAKAADNFTSILKOPALASGLTALGNV 456
434 VLSAPSLQDPPQALLMEAGTSLKTSDD-LKATLTSIPHSLOTEKSVYIHA-----N 486
Db 457 ELDVNGFTQTEGSTLLMQPGTKLADTEALSTKLVVDLSLAEKNKSVSETAGANKTIT 516
487 LSIQKIFLNSGDENFEYENVELLSKEQNNIPLTSLKEQSHLHPDNLSS-----HF 539
Db 517 LITSLIVQDSSG--NFYESHTINQATQPLVFTFAAADAIDYI-DALLTSPQVOTPRHY 573
540 GYQGDWTFESKMDSDGSHLIANMTPKNYVPHPEROSTIVANTLMNTYSDMQAOSMINTI 599
Db 574 GYQGHWEATWADISTAKSGMTWTTGTYNPNPERASVYVDSLWASTDIRTLOOIMTSQ 633
600 AHGAVLYFGTWSAVSNLFYANDSGKPIDNMHRSGLYLFISTSLDDHISFCLAQOL 659
Db 634 ANSLYQORGMASTANFPH-KDKSGTN-QAFRKSYGVTVGSAEDPSENTEVAFACQL 691
660 LKSSDSFTSTETTSYIATVOAQ-----LATSLMK-----ISAQACYNES 700
Db 692 FGKQKDLFIVENTSHNYLASLYLQHRAPLGGLPMPFSGITDMKIDPLLNQLSYST 751
701 IHELKTYSRFSKRGFGSMHVAASGEVCASTPI-VNSGCLFSSFSIFSKLQGFSTQD 759
Db 752 KNDMDITRYTSY-DEAQSMTNNSGALGLSLALYLTPKPAFFQGYPPFLKFGQVVSROQ 810
OY 760 GFEESGEIRFSFASFRNISLPIGITFEKKSKQKTRFYVFLAYIODLKRQVSGVYL 819
```


Oy 155 FSLQHNVLFTAFEEENS-KGNGCAIOAOTFSLSRVSP--ISFARRADLNGALCCSNL 211
 Db 179 FLISCTSOFAFSRQAFTKOGGVYATGTTIENSPOIVSFSONLAKSGGALYSTD- 237
 Oy 212 ICSGVN-PLFTFGNSA-----TNGAICCIDLNTSEKGSLSLACNOETTFASNSAKEK 265
 Db 238 NCSITDNOFVIFDGSNAWMAOAGCAIC-----TTTDR-TYTLGNKNLSSTNNLTALTY 292
 Oy 266 GCATYAKHNVLRNGPVSPFINNSAKI-----GCAIAIOGGSLSILAGEGVLFQNNQ 319
 Db 293 GCATISGLKVSISAGGPTLFQSNISSSGAGGGAINIAGELALATSODITF-NNNQ 351
 Oy 320 RTSDOGLVNAIYLEKDALISSLEARNG-DLFFDPPIVOESSKSPPLSSLOASTVST 378
 Db 352 VTNGSTSRNAINIITDAKVTSTIRATGOSIYFYDPTNPGTAASDTLNLNLADANS-- 409
 Oy 379 PATASPLVITQTSANRSVIFSSERLSEEEKT-PDNLTSOLOPTELKSGRLVKDRAYLSA 437
 Db 410 -----ELEYGCAIYFSGEKLSPTEKAIAANVTSTIRQAVLARGLVLRDGYTTF 460
 Oy 438 PSLSODPOLALIMEAGTSLKT--SSDLKIATLSTPLHSIDTEKSVTHA---PNLSIOKI 492
 Db 461 KDLTOSPGRIIMLDGGTTLTSAKEANLSLNGLAVALNSLDGTNKALKTEADKNISLST 520
 Oy 493 FLNSGDENFYENVELLSKEONNIPLLTSLKESQSHLDPGNS-----SHFGYQGD 544
 Db 521 IALIDTEGSEFYENHNL--KSASTYPLLELTLAGANTITLGAISTLTLQEPETHYGOYN 578
 Oy 545 WFFSKDSEGHSLIANMTPKNVVPHPEROSTLVANTLNNTYSDMOAVOSMINTIAHGA 604
 Db 579 WOLSMANATSSKIGSINMTRTGITPSPERKSNPLNSLGNFTIDISINQLIETKSSGEP 638
 Oy 605 YLEFGTWSAVSNLFYAHDSGKPIDMWHRSGLYLFGLISTHSLDHSFCLAAQOLLGKSS 664
 Db 639 FERELWLSGIANFFYRDSMPT--HGFRHISGVALGITTATPAEDQLTFACQLFARDR 696
 Oy 665 DSFTTSTETTSYIATVQAOALATSLKIS-----AQAACIN 698
 Db 697 NHITGKNHGDVTGASLYFHTTEGLFDIANFLWCKATRAPVWLSEISQIIPLSFDAFSLV 756
 Oy 699 ESTHELKTY--RSFKEGFGSHSAVSGEVCASIPYVNSGSLFSSFSIFSKLQFSG 756
 Db 757 HTDNHMKTYTTDNSITK---GSMRNDAPFCADLGASLPFYISVYLLKEVEPFYKQYIYA 813
 Oy 757 TODGFEESGEIRSFASFSFRNISLPITGITEFKKSOKTRTYVFLAYIIDLKRDVSGP 816
 Db 814 HQDEFERVAEGRAFNKSELINVEIPIGVTFERDSKSEKGYDILMLYILDVARRPKQ 873
 Oy 817 VLLKNAVSDAMANDSAIYFRLTNORALH-RLQTLNVCVLRGOSHSLDLGTT 875
 Db 874 TSLIASDANMAAGTNILAROGFSVRAANHPQVPHMEIFQGFAYEVYSSSRNNTNLGSK 933
 Oy 876 YRF 878
 Db 934 FCF 936
 RESULT 14
 B01591
 Polymorphic membrane protein G family CP0308 (Imported) - Chlamydomonas pneumoniae (str)
 C:Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae
 C:Date: 31-Mar-2000 #sequence-revision 31-Mar-2000 #text-change 11-May-2000
 C:Accession: B01591
 R:Read: T.D.: Brunham, R.C.: Shen, C.: Gill, S.R.: Heideberg, J.F.: White, O.: Hickey,
 C.: Dodson, R.: Gilm, M.: Nelson, W.: Deboy, R.: Kolonay, J.: McClarty, G.: Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis M09n and Chlamydia pneumoniae AR39.
 A:Reference number: AB1500; M01D:20150255
 A:Accession: B01591
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-936 <REA>

A:Cross-references: GB:AE002193; GB:AE002161; MID:q7189234; PIDN:AF38165.1; PID:q718
 A:Experimental source: strain AR39, HL cells
 A:Genetics:
 A:Gene: CP0308
 Query Match 17.0%; Score 772; DB 2; Length 936;
 Best Local Similarity 28.0%; Pred. No. 9e-41;
 Matches 236; Conservative 143; Mismatches 378; Indels 86; Gaps 22;
 Oy 97 DLISIONFRFISTDDSSKESPPS--IHOKNQLSLRNNGSMFCRNHAEGGCAISADA 154
 Db 119 NLLFNDPSRLSTIISCPILSLPTGOCALKSVGNLSLGNQIIFTFONPSDNGYINTKN 178
 Oy 155 FSLQHNVLFTAFEEENS-KGNGCAIOAOTFSLSRVSP--ISFARRADLNGALCCSNL 211
 Db 179 FLISCTSOFAFSRQAFTKOGGVYATGTTIENSPOIVSFSONLAKSGGALYSTD- 237
 Oy 212 ICSGVN-PLFTFGNSA-----TNGAICCIDLNTSEKGSLSLACNOETTFASNSAKEK 265
 Db 238 NCSITDNOFVIFDGSNAWMAOAGCAIC-----TTTDR-TYTLGNKNLSSTNNLTALTY 292
 Oy 266 GCATYAKHNVLRNGPVSPFINNSAKI-----GCAIAIOGGSLSILAGEGVLFQNNQ 319
 Db 293 GCATISGLKVSISAGGPTLFQSNISSSGAGGGAINIAGELALATSODITF-NNNQ 351
 Oy 320 RTSDOGLVNAIYLEKDALISSLEARNG-DLFFDPPIVOESSKSPPLSSLOASTVST 378
 Db 352 VTNGSTSRNAINIITDAKVTSTIRATGOSIYFYDPTNPGTAASDTLNLNLADANS-- 409
 Oy 379 PATASPLVITQTSANRSVIFSSERLSEEEKT-PDNLTSOLOPTELKSGRLVKDRAYLSA 437
 Db 410 -----ELEYGCAIYFSGEKLSPTEKAIAANVTSTIRQAVLARGLVLRDGYTTF 460
 Oy 438 PSLSODPOLALIMEAGTSLKT--SSDLKIATLSTPLHSIDTEKSVTHA---PNLSIOKI 492
 Db 461 KDLTOSPGRIIMLDGGTTLTSAKEANLSLNGLAVALNSLDGTNKALKTEADKNISLST 520
 Oy 493 FLNSGDENFYENVELLSKEONNIPLLTSLKESQSHLDPGNS-----SHFGYQGD 544
 Db 521 IALIDTEGSEFYENHNL--KSASTYPLLELTLAGANTITLGAISTLTLQEPETHYGOYN 578
 Oy 545 WFFSKDSEGHSLIANMTPKNVVPHPEROSTLVANTLNNTYSDMOAVOSMINTIAHGA 604
 Db 579 WOLSMANATSSKIGSINMTRTGITPSPERKSNPLNSLGNFTIDISINQLIETKSSGEP 638
 Oy 605 YLEFGTWSAVSNLFYAHDSGKPIDMWHRSGLYLFGLISTHSLDHSFCLAAQOLLGKSS 664
 Db 639 FERELWLSGIANFFYRDSMPT--HGFRHISGVALGITTATPAEDQLTFACQLFARDR 696
 Oy 665 DSFTTSTETTSYIATVQAOALATSLKIS-----AQAACIN 698
 Db 697 NHITGKNHGDVTGASLYFHTTEGLFDIANFLWCKATRAPVWLSEISQIIPLSFDAFSLV 756
 Oy 699 ESTHELKTY--RSFKEGFGSHSAVSGEVCASIPYVNSGSLFSSFSIFSKLQFSG 756
 Db 757 HTDNHMKTYTTDNSITK---GSMRNDAPFCADLGASLPFYISVYLLKEVEPFYKQYIYA 813
 Oy 757 TODGFEESGEIRSFASFSFRNISLPITGITEFKKSOKTRTYVFLAYIIDLKRDVSGP 816
 Db 814 HQDEFERVAEGRAFNKSELINVEIPIGVTFERDSKSEKGYDILMLYILDVARRPKQ 873
 Oy 817 VLLKNAVSDAMANDSAIYFRLTNORALH-RLQTLNVCVLRGOSHSLDLGTT 875
 Db 874 TSLIASDANMAAGTNILAROGFSVRAANHPQVPHMEIFQGFAYEVYSSSRNNTNLGSK 933
 Oy 876 YRF 878
 Db 934 FCF 936
 RESULT 15
 H01722


```

Db 506 -----NGNSTLYONTTIEOGRVLEKAKLVNSLSUTGGS-LYMEAGSTLDFV 553
QY 457 -----KTSDDLKATLSTPLHSI-----DREKSVTI 482
Db 554 TTPROPQOPRANQOLITLISLHLSLSSLIANNAYTPPTNPAPADSHPAIIGSTTACSVTI 613
QY 483 HAPNLSTOKIFLSNGDENFENVELLSKEON-----NIPULTLSKEQSHLILPGQNLSSH 538
Db 614 SGP-----IFEDLI-DOTAVDRYDMLGSNOKIDVLKIQLOLQPSANASDITLIGENMPK 666
QY 539 FCYGGDMTFW--KQSDG-HSLIANMTPKNVVPHPEPOSTLVANTLNNYSDMOAVOSM 595
Db 667 YGQGSWMLADPNTANNQPTLKATWTKTGYNPCPERVASILVPSLNGSLIDDISAISA 726
QY 596 INTIAHGAYLFTGWSAVSNLFVAHDSGKPIDMHHRSILCYLFGISTHSLDDHSPCLA 655
Db 727 IQASVDGSHYCGGLWVSVSNFFY-HDRDALG-QGYRYISGYSIGANSY-FGSSMFGIA 783
QY 656 AGOLIGKSSDSFTTST-----TTSYIATVQALATSLM-KISAQACYNESHELTXYR 709
Db 784 FTEVFGSRKDYVVCRRSNHACIGSYLLSTKQALCGSYLFGDAFIRASTYGFQNMHTST- 842
QY 710 SFSSKFGSMHSAVAVSGEVCASIPIVNSGGLF-SFSIFSKLOGFSGTODGFEESGEL 768
Db 843 TFAESDVRMDNCLVGLGVLPIYITPSKYLNLDRFVQAEFSYADHESFTEEGDA 902
QY 769 SFSASFRNISLPIGITEFKKSOKTRTYVEFGAYIQDLKRDVESGPVLLKANAVSWDA 828
Db 903 RAFRSGHMLNLSVPYGVKRCDSSTHPNKYSFGAVICAYRTIISOTYTLTSHQETWT 962
QY 829 PMANDSRAYMRLTNORAL-HRLQTLNVCVLGQSHSYSLDGTYYRF 878
Db 963 DAFHLARHGVIKRGMYASLTSNIEYGHGRYEYRDTSGYGLSAGSKVRF 1013

```

RESULT 17

```

B72131
polymorphic outer membrane protein g family - Chlamydomophila pneumoniae (strain CWL029)
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: B72131
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: B72131
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-922 <ARN>
A:Cross-references: GB:AE001585; GB:AE001363; NID:94376235; PIDN:AND18163.1; PID:9437626
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: pmp_1

```

Query Match 16.3%; Score 740; DB 2; Length 922;

Best Local Similarity 27.2%; Pred. No. 9.3e-39;

Matches 226; Conservative 156; Mismatches 362; Indels 88; Gaps 25;

```

QY 103 FRFLSFTDCSKESPSTIHOQKQOLSLRNGGMSFPCRNIAEESGGATSDAFSLQHNLY 162
Db 123 FSTLSFIOSPGDIKEOGLYSKNA-LMLNNYVVRFEONOSKRTGKATISGANVTIYGNVD 181
QY 163 FTAFENSSKNGGCAIOAOTFSLSRNVSPISFARNRA-DLNGCAIC-SNLIJCSGNVPL 220
Db 182 SVSFYQAAIFGGAHSSGLOIAVNOAELIRFQNTAKNGSGALYDGDIDQNNVYL 241
QY 221 FTGNSAT----NGCAICGISDLNTSEKSL-SLACNOETLFPASNAKKEKGATIAKMHV 275
Db 242 FREHEALTTAIGKGAACCLPTSGSSFPVPDIVTFSDMKQLVFERNHISIMGGATYARKLS 301
QY 276 LRNGPVSFIN-----NSAIGGAIAIOSGGSILAGEGVLFQNNNSORTSDOGLVRNA 330

```

```

Db 302 ISSGGPTLFINNISYANSONLGAIAIDTGCGLSISAEKGTITEGNN--RTSLPFL--NG 357
QY 331 IYLEDALISSLEARNG-DILFPOPIVQESSKESRPLSSIOASTPTPTAPASPLVIQT 389
Db 358 IHLLONAFLKLOARNNGSIEFYDITSEAD-----GSTOLNNGDPK-----NK 402
QY 390 SANRSVFSERLSEEEKTPDNLTLOOPIELKSGRLVLDRAVLSAPSLSDOPALLI 449
Db 403 EYTGITLFSGEKSLAND--PRDFKSTIPQNNLSAGLYIVINDEGAETVYSKTFQSGSLV 460
QY 450 MEAGTSLKTS-SDLKATLSTPLHSIDTEKSVTHAPNLSTOKIFLSNGD-----ENFY 503
Db 461 IDLGTKLASKEDIAITGLAIDIDLSSSSTAAVYKANTANKQISVTDSIELISPTGNAY 520
QY 504 ENVELLSKEONNIPULTLSK-BQSHLHLPDGN---LSHFQGGDMTFWMSKDSDE--GHS 557
Db 521 EDLRM--RNSOTFPLLSLEPGAGSVTYAGDFLPVSPHYGQNMKIAMTGTGNKVBGF 578
QY 558 LIANMTPKNVVPHPEPOSTLVANTLNNYSDMOAVOSMINTIAHGAYLFTGWSAVSNL 617
Db 579 F---WDKINRYKRPKEGCLVNPILMGNAVDRSLMOVQETHASSLQDRGLMIDGIGNF 635
QY 618 FYAHDSGKPIDMHHRSILCYLFGISTHSLDDHSPCLAQOLIGKSSDSFTTSTSYI 677
Db 636 F--HYASASEDNIRYRHNSGVLVSNNEITPKHYTSMASFQLSRDKDYAVASNEEYRML 693
QY 678 ATVOAQLATSLMKISAQA-----CYNESHELTXYRS 710
Db 694 GSLLYQTTSLNIFRYASRNPNNVNGILSRPFLQNLIFHFLCAYGATMDMTIDIAN 753
QY 711 FSKGFGSMHSAVAVSGEVCASIPIVNSGGLF-SFSIFSKLOGFSGTODGFEESGEL 769
Db 754 FPMVK-NSKRNKMCALIECGSMPLVLFENGRLFOGAIIPMKIQLYAVAGDPRKETTADR 812
QY 770 SFSSASFRNISLPIGITEFKKSOKTRTYVEFGAYIQDLKRDVESGPVLLKANAVSWDA 829
Db 813 RFSNGSLTISVPLGIRFEKLLALSDOVLYDFSEFSYIPDIFRKPDSCEALVIGSMLVP 872
QY 830 MANDSRAYMRLTNORALHRLQTLN--NVSVLRLGQSHSYSLDGTYYRF 878
Db 873 AAHVSHARFVSGTGRYHFNDDYELLRCGSIEC--RPHARNYNINCGSKFRF 922

```

RESULT 18

```

F81539
polymorphic membrane protein G family CP0770 (imported) - Chlamydomophila pneumoniae (s
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: F81539
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gyll, S.R.; Heideberg, J.F.; White, O.; Hicke
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Morn and Chlamydia pneumoniae AR39
A:Reference number: AB1500; MUID:20150255
A:Accession: F81539
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-922 <REA>
A:Cross-references: GB:AE002237; GB:AE002161; NID:g7189684; PIDN:AAF38570.1; PID:g718
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0770

```

Query Match 16.3%; Score 739; DB 2; Length 922;

Best Local Similarity 27.2%; Pred. No. 1.1e-38;

Matches 226; Conservative 156; Mismatches 362; Indels 88; Gaps 25;

```

QY 103 FRFLSFTDCSKESPSTIHOQKQOLSLRNGGMSFPCRNIAEESGGATSDAFSLQHNLY 162
Db 123 FSTLSFIOSPGDIKEOGLYSKNA-LMLNNYVVRFEONOSKRTGKATISGANVTIYGNVD 181
QY 163 FTAFENSSKNGGCAIOAOTFSLSRNVSPISFARNRA-DLNGCAIC-SNLIJCSGNVPL 220

```


RESULT 20
 DB1594
 polymorphic membrane protein E/F family CP0283 (imported) - Chlamydia pneumoniae (str
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
 C:Accession: DB1594
 R:Read, T.D.: Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
 A:Reference number: AB1500; MUID:20150255
 A:Accession: DB1594
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-946 <RNA>
 A:Cross-references: GB:AE002189; GB:AE002161; NID:97189205; PIDN:AAF38141.1; PID:9718920
 C:Genetics:
 A:Experimental source: strain AR39, HL cells
 A:Gene: CP0283

Query Match 12.7%; Score 575.5; DB 2: Length 946;
 Best Local Similarity 23.4%; Pred. No. 2.4e-28;
 Matches 231; Conservative 151; Mismatches 427; Indels 179; Gaps 33;

OY 13 AALISSTAVLFGQDPLGETALLTKNPNHYVCTFFEDCTMESLFP-----ALCAHASQ 64
 DB 16 ALILGKTTILNATPLSD--YFDNQAOLTTLPPLIDTLNMTPYSHRATLFGVRDDTNO 73
 OY 65 DDPLVYLGNSYCF-----VSKLHITDPKEAL-----FKKGDLSIQ-NFRF 105
 DB 74 DIVLHQNSTIESWFENFSODGALSCSLAITNTKNOILFLNSFAIKRAGAMVNGNFDL 133
 OY 106 -----LSFTDCSSKE---SSPSIIHOKNGQLSRNNGSMFCRNHAEAGSGAI 150
 DB 134 SENHGSIIIFSGNLSPFNASNPADCTCGAVLCSKNVTIS-KNGCTAVFLINNAKSSGCAI 192
 OY 151 SADAFSLQHNLYLTAFEENSSKGN-GGAIOAQTFLSRNVPSPISFARNRADLNGAICCS 209
 DB 193 QAAIINIKDNTGCPCLFFNNAGTAGALFANACRIENNSQPIYFLFNNQSLGAIKRVHQ 252
 OY 210 NLICSGNVNPLPFTGNSA-----TNGATCICISDLNTEKSGSLACNOETLFSAN 260
 DB 253 ECLITKNTGSIYFNNFAMEADISANHSSGAIYICIS--CSIKDNPGLA-----FDNN 304
 OY 261 SAKKGAATYAKHMLVRYNGVPVFINNSAKIGAIQSGSLSLAGEGSLVLFONNSOR 320
 DB 305 TAARDGALCTOSLTIDQSGPVYFTNNOGTWGAIMLRDGACTLPADODDIFYYNN-RH 363
 OY 321 TSDOGLVRNAYLEKDALISLEARNGDILFDPPI-----VOESSKESPLPSLQASVT 375
 DB 364 FKDTFSNVHVSVCNTRNVSITVGASOGHSATFYDPILORYTIONSIOKFNPNPEHL----- 418
 OY 376 SPPPATASPLVIOTISANRSVTFSSERLSEBEKPPDNTSQLOQPIELKSGRLVLDKRAVL 435
 DB 419 -----GTLFSSAVIPIOTISRDPIFSHRHHIIGLYNTALIEDRAEW 461
 OY 436 SAPLSLQDPOALLI-----MEAGSLKTSDDLKATLSIPLSL----- 474
 DB 462 KYVKPQDQGTLLGLSGRAVFTTIDEESSSSVGSVININMLAINLPSILGNRAVAKMIR 521
 OY 475 -----DTEKSVTIHAAPNLSTOKIFLNSGDENF--YENVLLSKRONNIPIL-TL 521
 DB 522 PTGSSAPYSDDNPIINLSP-----LSLDDENLDYDTRAD-LAOPIAEVPLLYLL 572
 OY 522 SKROSHL-----LPDG-NLSHFGYQGDWTFPSKMD-----SPEG-----HSLIANWT 563
 DB 573 DVYAKHINTNIFYPEGINTTOHYGYGVWSPYLETITTSQTSSEDTVNTLHROLYQDWT 632
 OY 564 PKNVVPEROSTLVANTLWNTYSDM-----QAVOSMINTIAGGAYLFTGWSAVSNL 617
 DB 633 PTGKVPKPNKGDIALSAFMSQFNNLFTALRYOTQOQIAPTAGSEA-----TR 681

OY 618 FYAHDSCGKPIIDNMHHRSLGYLFGISTHSLDHSFLACQULGKSSDSFTTETTSYI 677
 DB 662 LFVHONSNDKAKGFHMEATGSLGTTSTNASHSGVNFSQLFSLYESH-SDNSVAHHT 740
 OY 678 ATVOALATSLMK-----ISAQACYNESIHELTKYRSFSGEGFSGWHSVAVGEVCSIP 733
 DB 741 TVVALQINNPMLQERFSTASLAYSVMNHKASGYSGKIQTEGKCYSTTLGALSCSL 800
 OY 734 IYSNNGSLFSSSTISKLDGFGCTODGPEESGCELRPSA-SFFNISLPDITFEKKSO 792
 DB 801 LQWRSRPL-HETPEIQAIVRSNOTAFQESGDKARKFSVHKPLVNLVPLSQSAMESK 858
 OY 793 -KTRTYVFLGAYIODLKRDESGPVLLKNAVSDAPMANLDSRAVYRLTNQALH-R 850
 DB 859 FLPTPYMTELAYQVLYOONEVNAVSSLESSQMLSTTLARNAIAFKGNQIFEPK 918
 OY 851 LQTLNVCVLRGQSHSYSLDGTYYRF 878
 DB 919 LSVPLDYQGSVSSSTTHYLHAGTTFKF 946

RESULT 21
 C72075
 polymorphic outer membrane protein e/f family - Chlamydia pneumoniae (strain CWL0
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
 C:Accession: C72075
 R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
 Nature Genet. 21, 385-389, 1999
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A72000; MUID:99206606
 A:Accession: C72075
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-946 <RNA>
 A:Cross-references: GB:AE001631; GB:AE001363; NID:94376750; PIDN:AAD18610.1; PID:9437
 C:Genetics:
 A:Experimental source: strain CWL029
 A:Gene: pmp_18

Query Match 12.7%; Score 574.5; DB 2: Length 946;
 Best Local Similarity 23.4%; Pred. No. 2.8e-28;
 Matches 231; Conservative 151; Mismatches 427; Indels 179; Gaps 33;

OY 13 AALISSTAVLFGQDPLGETALLTKNPNHYVCTFFEDCTMESLFP-----ALCAHASQ 64
 DB 16 ALILGKTTILNATPLSD--YFDNQAOLTTLPPLIDTLNMTPYSHRATLFGVRDDTNO 73
 OY 65 DDPLVYLGNSYCF-----VSKLHITDPKEAL-----FKKGDLSIQ-NFRF 105
 DB 74 DIVLHQNSTIESWFENFSODGALSCSLAITNTKNOILFLNSFAIKRAGAMVNGNFDL 133
 OY 106 -----LSFTDCSSKE---SSPSIIHOKNGQLSRNNGSMFCRNHAEAGSGAI 150
 DB 134 SENHGSIIIFSGNLSPFNASNPADCTCGAVLCSKNVTIS-KNGCTAVFLINNAKSSGCAI 192
 OY 151 SADAFSLQHNLYLTAFEENSSKGN-GGAIOAQTFLSRNVPSPISFARNRADLNGAICCS 209
 DB 193 QAAIINIKDNTGCPCLFFNNAGTAGALFANACRIENNSQPIYFLFNNQSLGAIKRVHQ 252
 OY 210 NLICSGNVNPLPFTGNSA-----TNGATCICISDLNTEKSGSLACNOETLFSAN 260
 DB 253 ECLITKNTGSIYFNNFAMEADISANHSSGAIYICIS--CSIKDNPGLA-----FDNN 304
 OY 261 SAKKGAATYAKHMLVRYNGVPVFINNSAKIGAIQSGSLSLAGEGSLVLFONNSOR 320
 DB 305 TAARDGALCTOSLTIDQSGPVYFTNNOGTWGAIMLRDGACTLPADODDIFYYNN-RH 363
 OY 321 TSDOGLVRNAYLEKDALISLEARNGDILFDPPI-----VOESSKESPLPSLQASVT 375
 DB 364 FKDTFSNVHVSVCNTRNVSITVGASOGHSATFYDPILORYTIONSIOKFNPNPEHL----- 418

Mon May 7 09:06:41 2001

Search completed: May 6, 2001, 19:26:18
Job time: 257 sec

us-09-677-752-4.rpr

OY 563 TPKNVPHPEROST-----LVANTLMNTYSDMA 591
DB 677 TEIKYISSNSKGLTQYRSSAGVNFNGVNGMNSFLKKGAKVNFKLKPNENNT-SKPLP 735
OY 592 VOSMINTIAHGCAYLFGTWGSAVSNLFYAHDSGKPID-NMHRSLGYLFGISTHSLDH 650
DB 736 IRLPLNITATGG-----GSVFEDIVANHSGRGALMKSEINISGANFTLNSHYRGDD 788
OY 651 SFCLAGOLLKSSDSFISTETTSYATVOALATSLMKISAQACYNESHELKTK-YR 709
DB 789 AFKI-----NKDLTINATNS-----NFSLQTKDDFYD 816
OY 710 SFSKEGFGSMHSAVSG 726
DB 817 GYARNAINSTYNISILG 833

RESULT 3
US-08-530-198-4
Sequence 4, Application US/08530198
Patent No. 5869065
GENERAL INFORMATION:
APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GEME III, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530,198
FILING DATE: 13-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERNSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: JWB-1186
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-530-198-4

Query Match 3.3%; Score 151.5; DB 2: Length 1477;
Best Local Similarity 18.9%; Pred. No. 0.00013;
Matches 151; Conservative 123; Mismatches 274; Indels 249; Gaps 36;

OY 73 NSCYMVSUKLHTDPRKALFKKSG-----DISTONFRFLSTDCSSKSS-PSIIHQ-- 123
DB 143 NGCVFLINPNCITIGDAIINTNGFTASTLDISNENKARNFTFEOTKDALAEIVNHGL 202
OY 124 -----KNGQSL-----RNNGSMSCFRNHAEGSGAIS-----ADAFSLQHNLYLFTA 165
DB 203 IYVGKGSVNLIGGKVKNEGVISV-----NGGSISLLAGQKITTSIDIIINPTITYSIAA 255

OY 166 FEENS-----SKG-----NGCAIQAOFFSLSRNVPISFARNADLNGCAIC 207
DB 256 PENEAVNIGDIFAKGNGINVRATIRNOGKLSADSVSKDNGNIVLSAKKEGAEIGVYS 315
OY 208 CSNLLCSGNVNPDPFTGNSAT-NGCAICISDLNTESEKSGSLACNOETLFRASNAKE-K 265
DB 316 AONQAKG--GKLMITGCKVTLTKGAVI--DLSGKEG-----ETYLQGDGERGCK 362
OY 266 GCAITAKHMLRYNCPVPSFINNSAKIGAIQSGSLSILAGESSVLPONNSQNTSDG 325
DB 363 NGIOLAKKTSLEK--STINVSKEKEGFATWG--DIALIDGINAQSGDIAKTGG 416
OY 326 LVRNA-----LYLEKDALISLEARNGDIILEFDPTVOESSKESPSLQASVTSPTPATA 382
DB 417 FVETSGHDLFIKDNALIVDAKER-----LDFEDVY--STNAEDPL-----453
OY 383 SPLVLIQTSANRSVIFPSERLSEEEKTPDNI.TSLOOPIELK-----SGRL-----427
DB 454 -----FNNMG INDEPFTGTEASDPKKNSELTTLNNTT ISNLYKANMTMN 499
OY 428 VYKRAVLSAPLSODPQALLIM-----EAGTSLKSSDLKLTLSIPHS---LDTEKSY 480
DB 500 ITASRLITVNSSINIGSNHLLHSKQKQGVQIDGDTTSKGNLTLYSGCWVDPVHNKI 559
OY 481 TIHAPRLSIOKIFLSNGDENFYE--NVELLSK-----EONNIPV-----L 519
DB 560 TLDQGLNLTAAVAFEGGNKARDAANKIYAOGVITTTGEGKDFRANVNLNGTQGL 619
OY 520 TSKESQSHLPLDGNLSHFYOGDWTFSS-----WKDSDEGSHL-----ANW 562
DB 620 NISSVNNL--THNLSCFINISGNITINOTTRKNTSYQTSHDSHMVNSALNLETGANF 676
OY 563 TPKNVPHPEROST-----LVANTLMNTYSDMA 591
DB 677 TEIKYISSNSKGLTQYRSSAGVNFNGVNGMNSFLKKGAKVNFKLKPNENNT-SKPLP 735
OY 592 VOSMINTIAHGCAYLFGTWGSAVSNLFYAHDSGKPID-NMHRSLGYLFGISTHSLDH 650
DB 736 IRLPLNITATGG-----GSVFEDIVANHSGRGALMKSEINISGANFTLNSHYRGDD 788
OY 651 SFCLAGOLLKSSDSFISTETTSYATVOALATSLMKISAQACYNESHELKTK-YR 709
DB 789 AFKI-----NKDLTINATNS-----NFSLQTKDDFYD 816
OY 710 SFSKEGFGSMHSAVSG 726
DB 817 GYARNAINSTYNISILG 833

RESULT 4
US-08-469-880-4
Sequence 4, Application US/08469880
Patent No. 5876733
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:


```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentlin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO.: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1338 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-9

Query Match 3.2%; Score 144; DB 2; Length 1338;
Best Local Similarity 19.7%; Pred. No. 0.00056;
Matches 155; Conservative 131; Mismatches 314; Indels 188; Gaps 38

0Y 72 GNSYCFPSKLIITDPKRELFKEKDDSTONRFLFTDCCSSKSSPSLIHKNGLSLR 131
Db 403 GRTY-WNNTLVNTS-----GSKFSLSDS-----TGSSTGSPSI---RNLNLI 444
0Y 132 NNGSMSCFNHAEGSGG--AISADAFSLQHNYLETFAFEENSSKGNGAIOAOTFSLSRV 189
Db 445 TPNKATF--NIQGSITANSTIKASTMPKSMANYALFNDIDISVSGGGSVNFKNASSSN 502
0Y 190 SP--ISFAHRNRDLNGAICCSNLICSGVNPLE-----FTGNSAT----- 228
Db 503 QTPGVATIKQNPNVSGSGL--NLKRGSTETAFSTIENDLNLTNATGNTITIROVEGTDSR 560
0Y 229 -----NGCATCICSDLNTSE-KGSLSLACN-QETLFAANSKAKKGCATYAKH 273
Db 561 VNKGVAARKNITFEKGQINIFGSGKATTEIKGVTIKRNTNALRGANFAENKSPNLINAG 620
0Y 274 MYLRVNGPSPFINSNAKIGGAIAIQSGSLSLAG-----EGSVLPQNNNS----- 318
Db 623 VI--NNGCNLTITAGSTININIGNLTVSGANLQAITNTFNVAGS--FDNMGASNISTARCG 676
0Y 319 -----QRTSDQGLVNAATYLEKDAITLSSLEARNQDILEFD-----PIVOESSKESP 365
Db 677 AKFKDINNTSSLNITNTSDTTRITIKGINSKNSGDLNIDDKSDAEIQIGMISQEGN 736
0Y 366 I-P--SSLOASVTSTPATASPLVIQTSANNSVFFSSERLSFEFTPTDNLTSQLOQPLEK 423
Db 737 LTRISSQKVNITN-----QITIKAGVEGGRSDSSEAEVA--NLTIQTK--ELKL 780
0Y 424 SGRLVTK--DRAVLSAPSLQDPQALLIMEAGTSLKTSDDLKAT-----LSIPLHS 473
Db 781 AGDLNLSIGFNKAKITAKNGD-----LTIGNASGGMADAKKVTPEKVDSKISTDGNH 833
0Y 474 LDTEKSVTHAPNLSTOKIFLNSGDENFYENVELLSK--ONNIPLLTSLKEQSHLH 530
Db 834 VTUNSEV--KTSNGS-----SNAGNDN-STGLTILSAKDVTVNNV---TSHKITINISA 880
0Y 531 PDGNNLSHFQYQ-GDMTFSKKDSDIEGSLIANMTPKRYVPHPEROSTLVANTLNTNTSDM 589

```

```

Db      881 AAGVNTKEGTTIATGTCVSVEVTANQNGTINCKNTISQNVATATPENLVTTETAVALINATSGT 940
OY      590 QAVOSMINTIAHGAVLFGTWG---SAVSNLFYAHDSSGKPIDNMHRRSLGLEIGSTH 645
Db      941 VNISPKTODINGG---IESTSGNVNITASGNTLKVSNITGQDV-----TV 982
OY      646 SLDDHSFCLAGCOLLGKSSDSFITTSTETSTIATVQAL-ATSLMKISAQCYMESIH 704
Db      983 TADGALTTTGTSTISATGTGANITTKTGIDNGKVESSGSVTLVATGATLAV----- 1035
OY      705 KTKYRSPFSEKFGSHSAV--SGEVCASPIYNSGSLPSSPSIFSKLOG-PSGTODG 760
Db      1036 -----GNISQNVYTTADSGKLTSTVGSTINGTINSVTTSSOSDIDGTTISQNTVN 1085
OY      761 FEESSGEI 768
Db      1086 VTASTGDL 1093

RESULT 9
US-08-038-682-2
; Sequence 2, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILIUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Maltare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038.682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ. ID NO.: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-038-682-2

Query Match 3.0%; Score 134.5; DB 1: Length 1536;
Best local Similarity 19.4%; Pred. No. 0.0058;
Matches 153; Conservative 123; Mismatches 253; Indels 259; Gaps 38;

OY      120 ITIHKNGOLSLRNNGSMSCFCRNHAESGGAISADAFSLQHNILFTFAFEENSXKNGGAIQ 179
Db      76 VVIH---GRTMVGVDGKKKTIIRNSVD--AIIINKKOFINIDONENWQFLOEN---NNSAVF 125
OY      180 AOTFSLSRNNSPISFARNRNDLNGAICCSNLIICSGENVNPLFTFGSATNGAICICISDL 239

```

Db 554 WDVHKNISLGAQGNINITYAKODIAFEKGSNOVITIGOGTTISGNOKGFRFNVSJLNGTGS 613
 QY 665 DSFTSTETTSYIATVOAQLATSL--MKISAQACYNESIHLELTKYRSFSKEGFGSMH- 720
 Db 614 GLOFTTKRKNKYAITNKKEGTINISGVNISMVLPKNEGYD-KFKGRY-----WNL 665
 QY 721 --SVAVSGEVCASIPYVNSGGLFSSFSIESKLOGFSGTODFEBSGCEITSFSASSFR 777
 Db 666 TSLNVESEGEF--NLIDSRSDSAGTLTOPYNLNGISFNKD---TTFVVERNARVNF- 718
 QY 778 NISLPICI 785
 Db 719 DIKAPICI 726

RESULT 11
 US-08-530-198-2
 : Sequence 2, Application US/08530198
 : Patent No. 5869065
 : GENERAL INFORMATION:
 : APPLICANT: BARENKAMP, STEPHEN J
 : APPLICANT: ST. GEME III, JOSEPH W
 : TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
 : TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
 : NUMBER OF SEQUENCES: 8
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Shoemaker and Mattare, Ltd
 : STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 : STREET: Bldg. 1
 : CITY: Arlington
 : STATE: Virginia
 : COUNTRY: U.S.A.
 : ZIP: 22202-0286
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/530,198
 : FILING DATE: 13-DEC-1995
 : CLASSIFICATION: 424
 : ATTORNEY/AGENT INFORMATION:
 : NAME: BERKSTRESSER, JERRY W
 : REGISTRATION NUMBER: 22,651
 : REFERENCE/DOCKET NUMBER: JWB-1186
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (703) 415-0810
 : TELEFAX: (703) 415-0813
 : INFORMATION FOR SEQ ID NO: 2:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1536 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-08-530-198-2

Query Match 3.08; Score 134.5; DB 2; Length 1536;
 Best Local Similarity 19.4%; Pred. No. 0.0056;
 Matches 153; Conservative 123; Mismatches 253; Indels 259; Gaps 38;

QY 120 IIMKNGQSLRNHNSPCFNRHAECSGCAISADAFSLHNYLTFAPFEENSSKNGCAIQ 179
 Db 76 VVH---GTATMOYDGNKNTITIRNSVD---AIIIMKOFINDONENKQFIQEN---NNSAVF 125
 QY 180 AOFISLSRNPSPISFARNRADLNGAICCSNLCISGVNVPFLFTGNSATNGAICISDL 239
 Db 126 NRTYS-----NOISOLKGLIDSNQOVFL-----INP-----NGITIGKQALI 162
 QY 240 NTEKGSLSLACNOETLFASN---SAKEKGALIAKH---MVLRYNGPVSEFINNSAKIGG 293

Db 163 NTNGTASTLSDLSNENIKARNFTFEQTKOKALAEIVNHGLIIVGKDGVSJLIGKVKNEG 222
 QY 294 AIAIOGSGSLSLIAGEBSVLFONNSORTSDGLVNRNAIL-----EKDAI-LSSLFARG 347
 Db 223 VISV--NGGSIISLAG-----OKITISDIINPTIYISAABENEAIVNLGDIFAKGG 271
 QY 348 DILFDPRIYQESSKESPLPSSLOASVTSPTPATNSPLVIQTSANRSVYFSSERLSEEEK 407
 Db 272 NI-----NVRATIRNKGKLSADSV 291
 QY 408 TPDNLTSLOQPIELKSGRLVLEKDR-----AVLSAPLSODPOLLIIEAGTSIKTSS 460
 Db 292 SKD-----KSGNIVLSAKEGEAEIGGIVISAQNOAKGGLMTGKVKLTGTA 339
 QY 461 DLKLA-----TLSTPL-HSLDEKSVTHAD----- 485
 Db 340 VIDLSKEGGETYLGDEREGKNGIQLAKTSLSEKSTINVSKEKGGRAIWDGIALI 399
 QY 486 --NLSTOKI-----FLNSGDENFENVELLSKEQ-----NNIPLTLKESQSHL 530
 Db 400 DGNINAGSGDIAKYGFVETSGHDLFIKDNAIVDAKEWLLDFDNVSIHAETAGRS----- 455
 QY 531 PGNLSHFEGYGDWTFESWKSDSEHSLIANMTPKNYVPHPEROSTLVANTLM-----N 584
 Db 456 ---NTSEDEDTG-----SGNSAS---TPKR-----NKEKTLNTLTLESILKKG 494
 QY 585 TYSDMAA-----VOSMINTAMGAYLF--GTWGSVAV---SNLFYHDS-----SK 626
 Db 495 TFEVITANQRIYVNSSIN-LSNGSLTLMSEGRGGVEIINDITTDGDTGANTLIYSG 553
 QY 627 PIDNMHHRSLGYLFGISTHSLDHSFLAAGOLL-----GKSS 664
 Db 554 WDVHKNISLGAQGNINITYAKODIAFEKGSNOVITIGOGTTISGNOKGFRFNVSJLNGTGS 613
 QY 665 DSFTSTETTSYIATVOAQLATSL--MKISAQACYNESIHLELTKYRSFSKEGFGSMH- 720
 Db 614 GLOFTTKRKNKYAITNKKEGTINISGVNISMVLPKNEGYD-KFKGRY-----WNL 665
 QY 721 --SVAVSGEVCASIPYVNSGGLFSSFSIESKLOGFSGTODFEBSGCEITSFSASSFR 777
 Db 666 TSLNVESEGEF--NLIDSRSDSAGTLTOPYNLNGISFNKD---TTFVVERNARVNF- 718
 QY 778 NISLPICI 785
 Db 719 DIKAPICI 726

RESULT 12
 US-08-469-880-2
 : Sequence 2, Application US/08469880
 : Patent No. 5876733
 : GENERAL INFORMATION:
 : APPLICANT: Barenkamp, Stephen J.
 : TITLE OF INVENTION: High Molecular Weight Surface Proteins
 : TITLE OF INVENTION: OF NO. 5876733-Typeable Haemophilus
 : NUMBER OF SEQUENCES: 8
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Shoemaker and Mattare, Ltd.
 : STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 : STREET: Bldg. 1
 : CITY: Arlington
 : STATE: Virginia
 : COUNTRY: U.S.A.
 : ZIP: 22202-0286
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/469,880
 : FILING DATE: 06-JUN-1995


```

Db 76 VVH---GTAATMOVDGNKTTIIRNSVD---AIIWKKOPNIDQENVOFLQEN---NNSAVF 125
Oy 180 AQTSLSRNVSPISFARNRADLNGAICCSNLICSGVNPPLFFTGNSATNGAICISDL 239
Db 126 NRVTS-----NOLSQLGLDLSNGQVFL-----INP-----NGITIGDAII 162
Oy 240 NTSEKGSLSIACNOETLPAASN---SAKEKGAIIYAKH--MWLRNGPVSFINSKICG 293
Db 163 NTNGFTASTLDISNENIKARNFTFEOTKDALAEIVNHGLITVGKGSVNLIGKVKYKNG 222
Oy 294 AIIAIOGSGSLIAGCGSVLPFNNSORTSDQGLVRNAIYL-----EKDAI-LSSLEARNG 347
Db 223 VLSV-NGGSISLLAG-----OKITISDIINPTIITYSIAAPENEAVALDIPAKGG 271
Oy 348 DILFFPIQVQESSKESPLSSLSQASVTSPTPATASPLVYIOTSANRSVFSERLSEEEK 407
Db 272 NT-----NVRRAATIRNGKLSADSV 291
Oy 408 TPDNLTSQLQPIELKSGRLVLKDR-----AVLSAPSLSDPPQALLIMEAGTSIKTSS 460
Db 292 SKD-----KSGNIVLSAKEGEAEIGVISAQNOQAKGKLMITGDKVTLKTGA 339
Oy 461 DKLAL-----TLSIPL-HSLDTEKSVTIHAP-----485
Db 340 VIDLSGKEGEGEYLOGDEREGCKNGIOLAKKTSLKCGSTINVSCKEKGRAIWMGDIALI 399
Oy 486 --NLSTOKI-----FLNSGDENFEYENVLLSKEQ-----NNIPLTLSEKQSHLH 530
Db 400 DGNINAGSGDLAKTGTFVETSGHDLFIKDNALYDAKEMWLDPDVNSINAEIAGR-----455
Oy 531 PGNLSSHFGYGDWTFWKKDSEGHSLANMTPKNVVPHPEROSTLVANTLW-----N 584
Db 456 --NTSEDEYTG-----SGNSAS--TPKR-----NKEKTLTLTLESILKKG 494
Oy 585 TYSDMA-----VOSMINTIAHGAVLYF--GTWGSAY--SNLFYAHDS-----SGK 626
Db 495 TPNVNTIARORIYVNSIN-LSNOSLTLWSEGSNGGVEIINDITGDDIRGANLTIYSGG 553
Oy 627 PIDNMHRSILGYLFGISTHSLDHSFLAAGQL-----GKSS 664
Db 554 WVDVHNINISLAGOGNINITAKODIAFEKGSNOVITGOGITTSNGOKGFRRNNSLWGTGS 613
Oy 665 DSFTSTSTSTYATVQAOIATSL--WKISAQACYNESIHELKTYRSEKGEFGCSWH- 720
Db 614 GLOFTTKRNTKAYAITKKEGTLINISGVNIMVLPKRNESGD-KFKGRY-----WNL 665
Oy 721 ---SVAVSEGCASIPVNSGSLFSSSFTSKLQGFSGTODGFESSSGEIRSFSSSFR 777
Db 666 TSLANSESEF--NLFTIDSGSDSAGTLTPYVNLNGISFNKD---TFVVERNARVNF- 718
Oy 778 NISLPIGI 785
Db 719 DIKAPIGI 726

```

```

RESULT 14
US-08-617-697-2
Sequence 2, Application US/08617697
Patent No. 5977336
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Matlare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Belklesser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-617-697-2

Query Match 3.0%; Score 134.5; DB 2; Length 1536;
Best Local Similarity 19.4%; Pred. No. 0.0058;
Matches 153; Conservative 123; Mismatches 253; Indels 259; Gaps 38;

Oy 120 ITHQKNGOISLRNNGSMSCRNHAEBSGGAISADAFSLQHNLTFTFEENSSKNGCAIQ 179
Db 76 VVH---GTAATMOVDGNKTTIIRNSVD---AIIWKKOPNIDQENVOFLQEN---NNSAVF 125
Oy 180 AQTSLSRNVSPISFARNRADLNGAICCSNLICSGVNPPLFFTGNSATNGAICISDL 239
Db 126 NRVTS-----NOLSQLGLDLSNGQVFL-----INP-----NGITIGDAII 162
Oy 240 NTSEKGSLSIACNOETLPAASN---SAKEKGAIIYAKH--MWLRNGPVSFINSKICG 293
Db 163 NTNGFTASTLDISNENIKARNFTFEOTKDALAEIVNHGLITVGKGSVNLIGKVKYKNG 222
Oy 294 AIIAIOGSGSLIAGCGSVLPFNNSORTSDQGLVRNAIYL-----EKDAI-LSSLEARNG 347
Db 223 VLSV-NGGSISLLAG-----OKITISDIINPTIITYSIAAPENEAVALDIPAKGG 271
Oy 348 DILFFPIQVQESSKESPLSSLSQASVTSPTPATASPLVYIOTSANRSVFSERLSEEEK 407
Db 272 NT-----NVRRAATIRNGKLSADSV 291
Oy 408 TPDNLTSQLQPIELKSGRLVLKDR-----AVLSAPSLSDPPQALLIMEAGTSIKTSS 460
Db 292 SKD-----KSGNIVLSAKEGEAEIGVISAQNOQAKGKLMITGDKVTLKTGA 339
Oy 461 DKLAL-----TLSIPL-HSLDTEKSVTIHAP-----485
Db 340 VIDLSGKEGEGEYLOGDEREGCKNGIOLAKKTSLKCGSTINVSCKEKGRAIWMGDIALI 399
Oy 486 --NLSTOKI-----FLNSGDENFEYENVLLSKEQ-----NNIPLTLSEKQSHLH 530
Db 400 DGNINAGSGDLAKTGTFVETSGHDLFIKDNALYDAKEMWLDPDVNSINAEIAGR-----455
Oy 531 PGNLSSHFGYGDWTFWKKDSEGHSLANMTPKNVVPHPEROSTLVANTLW-----N 584
Db 456 --NTSEDEYTG-----SGNSAS--TPKR-----NKEKTLTLTLESILKKG 494
Oy 585 TYSDMA-----VOSMINTIAHGAVLYF--GTWGSAY--SNLFYAHDS-----SGK 626

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/212.133A
  FILING DATE: March 11, 1994
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/864,004
    FILING DATE: 07-APR-1992
  ATTORNEY/AGENT INFORMATION:
    NAME: Padst, Patrea L.
    REGISTRATION NUMBER: 31,284
  REFERENCE/DOCKET NUMBER: EMU/76677
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 404-572-6508
    TELEFAX: 404-572-6555
  INFORMATION FOR SEQ ID NO: 8:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 2319 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: protein
      HYPOTHETICAL: YES
      ANTI-SENSE: NO
      FRAGMENT TYPE: N-terminal
      ORIGINAL SOURCE:
        ORGANISM: Mus musculus
        PUBLIC INFORMATION:
          AUTHORS: Eider, F.
          AUTHORS: Gitschler, J.
          TITLE: Sequence of the Murine Factor VIII cDNA.
          Patent No. 5663060
          JOURNAL: Genomics
          VOLUME: 16
          PAGES: 374-379
          DATE: 1993
          RELEVANT RESIDUES IN SEQ ID NO: 8: FROM 1 TO 2319
US-08-212-133A-8
  
```

```

Query Match      2.9% Score 132.5: DB 1: Length 2319;
Best Local Similarity 20.2% Pred. No. 0.02;
Matches 151: Conservative 109; Mismatches 300; Indels 187; Gaps 35;

QY 46 FEDCT-----MELPAPL-----CANASODPLVIGNSYCWVSKLHTDPKALF 92
DB 775 FDDSTIPKMDMEIEPOFEIEIAEMLKQVSYSVSDMLMLGQSHR-TPHGLFLSDGGEALY 833
QY 93 K-----EKGDLSIONFRLFTDCSSKSSPSI1HOKNGQSLRNNGSMFC 139
DB 834 EAIHDDSPNAIDSNCPKVTQLR-----PESHSEKIVFTPOPGLRNSKSL--- 883
QY 140 RNHAESSGAISDAISLQHNITFT-----AFEE-----NSKNGGAI 178
DB 884 ETTIEKMKKLGQVSSLPNSLMTTTLISDNLKATEKTDSSGCFPMVHSSSKLSTTAR 943
QY 179 QAOFTLSRNVSPIFARNRDLNGAICCSNLICSGNVNPLFFGTNSATNGAICISD 238
DB 944 GKRAYSLVGSHPVNLNSEEKSD-----SNLDS-----TLMYSGESLPRDILSTEND 991
QY 239 LNTSEK--GSLSLACNOETLFLASNSAKKEGAIY---AKHMYLRNCPVPSFINNSAKIGG 293
DB 992 RLRLKRFHIALTLKDNFTLFKDNVSLMKTKYNNHSTTNEKLHTSPSTIENSSTDLQD 1051
QY 294 AI-----AIOSGGSLILAGEGVSFLFONNSQRTSDQGLVNMALYLEDAISL-SLEARNG 347
DB 1052 AILKVNSEIO---EYVALIHDGTLCKNST-----YLRNLNHLNRTTSTKNK 1095
QY 348 DILF---FPIVQESSKESPLSSLOASVTSPTPATASPLVIQTSANRSVIFSSERLSE 404
  
```

```

DB 1096 DIFHRKDEDPIDPEENTIMPFSKMLFLS-----ESSMWEKKTNGNNSL-----NSE 1142
QY 405 EERTDNLTLSQOP1ELKSGRLVLDRAVLISAP---SLSDOPQALLIMEAGTSLSKTSDD 461
DB 1143 QHSPKPOLVLYL-----FKRYVKNOSFLEKKNVYVEODGFTKNGLKMDAPPHNNS 1194
QY 462 LKLTATSLPLHS---LDTEKSV--TIHAPNLSTQKIFLSN---SGDENFEYENVELLSKE 512
DB 1195 IFLTTIS-NVHENGKRNDEKNIOEIEKALIEKAVYLPOVHEATGSKNPLKDLILGTR 1253
QY 513 QN-----NIPLLTLSKEQSHLPLPDGNLSSHFQYOGDWTFSKSDSECHSLIA 560
DB 1254 QNISLEYEVHPVLQNTITSINNSTNTVOIHM-----EHFFKRR-----KDKETNSEGLV 1301
QY 561 NMTPKNYVHPEROSTLVANTIMNTYSDMQAVOSMINTIARHGAYLFGTWSAVSNLPTA 620
DB 1302 NKTRE-----MVKN-----YPSQKNITTRSKRALQOFRISTOW---LKTINCS 1342
QY 621 HDSSGKPIDNMHRSGLYFGISTHSLDHSFLCAGOLLGKSSDSFSTETTSYIATV 680
DB 1343 TOCIIKQID--HSEKMKR--ITKSSLSDSVYIKSTTQ--NSSDSHYKTSAPFPIDLK 1396
QY 681 QAOLATSLMKISAQACYNESHELTKR 707
DB 1397 RSPFQNKFSHYOA---SSYIYDFKTK 1419
  
```

```

RESULT 17
US-08-474-503-6
; Sequence 6, Application US/08474503
; Patent No. 5744446
; GENERAL INFORMATION:
  APPLICANT: Emory University
  TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
  NUMBER OF SEQUENCES: 12
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Kilpatrick & Cody
    STREET: 1100 Peachtree Street, Suite 2800
    CITY: Atlanta
    STATE: Georgia
    COUNTRY: US
    ZIP: 30309
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/474,503
      FILING DATE: 07-JUN-1995
      CLASSIFICATION: 435
      ATTORNEY/AGENT INFORMATION:
        NAME: Pratt, John S.
        REGISTRATION NUMBER: 29,476
        REFERENCE/DOCKET NUMBER: EMU106CIP(3)
        TELECOMMUNICATION INFORMATION:
          TELEPHONE: 404-815-6500
          TELEFAX: 404-815-6555
      INFORMATION FOR SEQ ID NO: 6:
        SEQUENCE CHARACTERISTICS:
          LENGTH: 2319 amino acids
          TYPE: amino acid
          STRANDEDNESS: single
          TOPOLOGY: linear
          MOLECULE TYPE: protein
          HYPOTHETICAL: YES
          ANTI-SENSE: NO
          FRAGMENT TYPE: N-terminal
          ORIGINAL SOURCE:
            ORGANISM: Mus musculus
            PUBLIC INFORMATION:
              AUTHORS: Eider, F.
              AUTHORS: Lakich, D.
  
```

```
QY 140 RNHAEGSGAISAQAFSLQHNVLFT-----AFEE-----NSSKNGCAI 178
DB 884 ETTIEVKMKKLGLOVSSLPNMTTILSDNLIKATEKTDSSGFPDMVHSSKLSITFAF 943
QY 179 QAOTFSLRNVPISFARRADLNGAICCSNLICSGVNNPLFTGNSATNGAICICSD 238
DB 944 GKKAISLVGSINPLANSSENSD-----SNILDS-----TLMYQESLPRONILSIEND 991
QY 239 LNTSEK--GSLSLACNOETLFPASNAKEKGAIY---AKHMYLRINGPVSFINSKAKIG 293
DB 992 RLREKRFHGIALTLTKDNTLFDKNVSLMKTNTKYNHSTNEKLHSTESPTSIENSTDLDD 1051
QY 294 AT-----AIOGSGSLILAGEGSLVFONNSQRTSDOGLVRNAILYEKDAIIS-SLEARN 347
DB 1052 AILKVNSEIQ--EYVALIHDGTLGKNST-----YLRNMLNRTSTKTK 1095
QY 348 DILF---FDPVIOESSKESPLPSLSQASVTSPTPATASPLVLOTANSRVSIFSERLSE 404
DB 1096 DIFHKDEDPIDODEENTIMPSKMLFLS-----ESSNMFKKTNGNNSL-----NSE 1142
QY 405 EKPDPNLTSQLQPIELKSGRLVLDRAVLAP---SLSDPQALLINEAGTSLKTSDD 461
DB 1143 QEHSPKQVLVYLM-----FKKYVKNOSFLSEKNKVTYEDGFTKNIGLKDMAFPHNMS 1194
QY 462 LKATLSTPLSL-----LDTEKSV--TIHAPNLSIOKIFLSN---SGDENFVENELLSKE 512
DB 1195 IFLTLLS-NVHENGHNQEKNIQOEIEKEKLEEVVLPOVHEATGSKNFKLDIILGTR 1253
QY 513 QN-----NIPLLTSKEQSHLPLPDGNTLSHFQYGDWTFMSKWDSEGHSLIA 560
DB 1254 QNISLVEYHVPYLOMTSITNNTSTNVQIHM-----EHFFKRR-----KQKETNSEGLV 1301
QY 561 NMTPKNVYHPERKOSTLVANTLMNTYSDMAVQSMINTIAGCAVLFGWGSVAVSLFA 620
DB 1302 NKTRE-----NVKN-----YPSQKNITTOKSKRALQOFRUSTOW--LKTINCS 1342
QY 621 HDSGKPLDNMHRSLGVLFGISTHSLDHSFCIACGLGKSSDSFSLTSTSTVIAV 680
DB 1343 TOCIITKOD--HSKEKKF--ITKSSLSDSVYIKSTTOT--NSSDSHIYKTSAPFPIDLK 1396
QY 681 QAOLATSLMKISAQACYNESIHLELTK 707
DB 1397 RSPFQNKFSHVQA-----SSYIYDFKTK 1419
```

```
RESULT 19
US-09-037-601-6
: Sequence 6, Application us/09037601
: Patent No. 6180371
: GENERAL INFORMATION:
: APPLICANT: Lollar, John S.
: TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
: NUMBER OF SEQUENCES: 40
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
: STREET: 5370 Manhattan Circle Suite 201
: CITY: Boulder
: STATE: Colorado
: COUNTRY: USA
: ZIP: 80303
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/037,601
: FILING DATE: 26-JUN-1996
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/US94/13200
: FILING DATE: 15-NOV-1994
: PRIOR APPLICATION DATA:
```

```
APPLICATION NUMBER: US 08/212,133
FILING DATE: 11-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 75-95F
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2319 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Mus musculus
PUBLICATION INFORMATION:
AUTHORS: Lakich, D.
AUTHORS: Gitschier, J.
TITLE: Sequence of the Murine Factor VIII cDNA
Patent No. 6180371
JOURNAL: Genomics
VOLUME: 16
PAGES: 374-379
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 1 TO 2319
US-09-037-601-6

Query Match 2.9% Score 132.5 DB 4: Length 2319:
Best Local Similarity 20.2% Pred. No. 0.02 Indels 187 Gaps 35:
Matches 151: Conservative 109: Mismatches 300:

QY 46 FEDCT-----MELFPAL-----CAHASODPLVYLGNSYCWFSKLTIDPREALF 92
DB 775 FKDSITPKNDMEKIEQFEIEAEMLKQVSYSVDMMLLGQSHP-TPHGLFLSDQGEALY 833
QY 93 K-----EKGLDSIONFRELSTDCSSKESPSIITHQNGQLSRNNGSMFC 139
DB 834 EAIHDDSPNAIDSMCEGPKVQQLR-----PESHSEKIVFTPOPGIQLRSNKSLL-- 883
QY 140 RNHAEGSGAISAQAFSLQHNVLFT-----AFEE-----NSSKNGCAI 178
DB 884 ETTIEVKMKKLGLOVSSLPNMTTILSDNLIKATEKTDSSGFPDMVHSSKLSITFAF 943
QY 179 QAOTFSLRNVPISFARRADLNGAICCSNLICSGVNNPLFTGNSATNGAICICSD 238
DB 944 GKKAISLVGSINPLANSSENSD-----SNILDS-----TLMYQESLPRONILSIEND 991
QY 239 LNTSEK--GSLSLACNOETLFPASNAKEKGAIY---AKHMYLRINGPVSFINSKAKIG 293
DB 992 RLREKRFHGIALTLTKDNTLFDKNVSLMKTNTKYNHSTNEKLHSTESPTSIENSTDLDD 1051
QY 294 AT-----AIOGSGSLILAGEGSLVFONNSQRTSDOGLVRNAILYEKDAIIS-SLEARN 347
DB 1052 AILKVNSEIQ--EYVALIHDGTLGKNST-----YLRNMLNRTSTKTK 1095
QY 348 DILF---FDPVIOESSKESPLPSLSQASVTSPTPATASPLVLOTANSRVSIFSERLSE 404
DB 1096 DIFHKDEDPIDODEENTIMPSKMLFLS-----ESSNMFKKTNGNNSL-----NSE 1142
QY 405 EKPDPNLTSQLQPIELKSGRLVLDRAVLAP---SLSDPQALLINEAGTSLKTSDD 461
DB 1143 QEHSPKQVLVYLM-----FKKYVKNOSFLSEKNKVTYEDGFTKNIGLKDMAFPHNMS 1194
```



```

QY 551 DSDEGSLIANTPKNYVHPEROSTLVANTLMTYSDMAVSMINTIAHGCAYLFCWT 610
Db 665 -EEG-----KN-----NGNLANTFKGSKSEONRF--LTTGTNLNGDL 698
QY 611 GSAVNLFYAHDSQKPIDNMHRSGLYFGISTSLDHSCLAGOLLGSSDSFTIS 670
Db 699 KVEKTELL-----SGRPT--HARDIA---GISSTKKDOH---FANNVVEDDWINN 746
QY 671 TETSYIATVOAO--ATSLMKISAO--ACYNESIHE-----LTKYRSF----- 711
Db 747 FKATINVTNNATLXSGRVANITSNITASDNKAVHIGKAGDTVCVRSDYGYVCTTD 806
QY 712 --SKGPGSMHVAASGEVCASIPYVNSGGLFSSFSJPSKLOGFSGODGFEES---- 765
Db 807 KLSDALNSFNATVNSGVNLS-----GNANFVLKANLFGTISGTNSOVRLTENSIMHL 862
QY 766 -GEIRSFASFRNISLPIG--ITFEKKSOQTRTY-----YFELGAYIIDL 808
Db 863 TGD-----SNVNOJLQKCHILHNAQNDANKVTYNTLTVNSLSONGSFY----YLTDL 912
QY 809 KRDVESPVYLLKANV--SWDAPMANLDSRAVYFRLTNOALHRLQTLNLVSCVLRGQSHS 867
Db 913 S--KOGDKVYVTKSATGNETLQVADKTGPTKNEILLFEDASHNATRNMLNVSIV-----G 965
QY 868 YSLDLC 873
Db 966 NTFVDLG 971

RESULT 24
US-09-356-952-4
: Sequence 4, Application US/09356952
: Patent No. 6117663
: GENERAL INFORMATION:
: APPLICANT: Borjack-Sjodin, Ann
: APPLICANT: Margarit, S. M.
: APPLICANT: Bor-Sogi, Dafna
: APPLICANT: Cole, Philip
: APPLICANT: Kuriyan, John
: TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
: FILE REFERENCE: 600-1-228N
: CURRENT APPLICATION NUMBER: US/09/356,952
: CURRENT FILING DATE: 1999-07-19
: EARLIER APPLICATION NUMBER: 60/093,631
: EARLIER FILING DATE: 1998-07-21
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 4
: LENGTH: 1589
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
US-09-356-952-4

```

```

Query Match 2.8%; Score 126; DB 4; Length 1589;
Best Local Similarity 18.1%; Pred. No. 0.04;
Matches 162; Conservative 141; Mismatches 322; Indels 272; Gaps 37;

```

```

QY 112 SSKESSPSIIHKNGLSLRNNGSMFCRNHAEGSGAITSADA-FSLQNHVLTFARENS 170
Db 143 SSASSSRSSLN-----SLGNSAYLVHPNPKPSRRGSSSTLSASLSNAH-----AETSS 191
QY 171 SKNGGAIOAOTFSLSRNVSPISFARNRADLNGAICCSNLICSGVNNPLFTGNSATNG 230
Db 192 GHANTVSMNNSPISAPNDASHIT--POSSNFNMSAISIQQMTKSAD-----GSSEMT 242
QY 231 GAICCSLDLTSKSGSLSLACNOETLFASNSAKEKGCALYAKHMV-----LRYNCP 281
Db 243 NAI-----NNNNE-----TNLOT-----SGEKAGPLVAEETIKILPLEIEIEMIING- 284
QY 282 VSRINNSAKIGALAIQSGSLSILAGEGSLVLPONNSQPRSDQIVRNALYLRKD----- 336
: : : : :
: : : : :

```

```

Db 285 --IRSNIASTWSPILIT-----KTSQKLV---YYNKDLIDYC 318
QY 337 ---AIISSLEARNIGDILFFDP-----IVQSSSKESPSPSLQASVTSPPPATAS 383
Db 319 SELPLISNIMESDIOCESEPFPPNDHLVNLVYTRDLRKNANIEDSSSTRKSESEONNS 378
QY 384 PLVIQTSANRVSIFSSERISEEKTDPNLTISOLOO--PELKSGLVLYKRAVLASAPIS 441
Db 379 SLMEKQDSKETDGNVNSINDDNNNNNKNEFNAGBSLNS-----LSAPDLT 428
QY 442 ODPQALLIMEAGTSKTS-----DLKATLSIPHSIDTEKSVYIHAPLSIQKIF 493
Db 429 QNIGSRVAVPSRSSLSIAKSDIFVYSRDILKMT---ELQDLVYVYTKTAH-----KMF 478
QY 494 LNSGDENYENVELLSKE-----QNNIPLITLSKE-----Q 525
Db 479 LKEN-RLNFTKRYFDLSDSIVFTOLGCRMLQHEIRAKSCSEIKRIFKGLISSLSRISIN 537
QY 526 SHLHLPDGNLSSHFGYQDWFPMKQSD-----EG-----HSLIA----- 560
Db 538 SHLY-----FDSAFIRKKMDTMNDKNDQENNCSTBEDGKIEVDSVHDLVSPYLSGK 592
QY 561 -----NMT-----PKNYVHPEROSTLVANTLMTYTS--- 587
Db 593 RNVTSTDTLTIPMRSSFVNENDMENFVGLPFRSVSVYTPRTSIGNSTLEDFSPSN 652
QY 588 ----DMQAVOSMINTIANGAILEFTWGSVSNLYAHDSQKPIDNMHRSGLYLF--G 641
Db 653 KNFKSAKSIYEMVDV-----EFSKFLRHVOLLVFLVQSVFSDNTLPLQDLPRFFKGS 705
QY 642 ISTHSDHSFCLANQOLLKSSDSFISTENT-----SYIATVOAOIATSLMKISAQ 695
Db 706 FSGGSWTNPSTFTTIDFEGNATKNKAVTSEVYASSKSSISIRPPKADAIASAGYS 765
QY 696 CYNESIHLELTK-----YRFSKEGFGSWHSAVGEVCASIPYVNSGGLF--SS 744
Db 766 ANSETNSOJDLKASSAASGVTFPRPNHRTFSARAKSKKKKPLVLDLNTMKKS 825
QY 745 FSTFSKLOGFCGTODGFEESGGEIRFSASFRNISLPIGITFEKKSOQTRTYVFLGAY 804
Db 826 SOIFEKLNATG-----EHLKIISKPSRIINLEIN--STYEIQINQ-----NVL 868
QY 805 IODIAKRDVESGPVYLLKANVSNAPMANLDSRAVYFRLTNOALHRLQTLNLVSCVL 861
Db 869 LLEILELDSLIFINLKNLIKTPSILLDESEFL-----VHAMSVSVYL 914

RESULT 25
US-08-687-956A-23
: Sequence 23, Application US/08687956A
: Patent No. 5861157
: GENERAL INFORMATION:
: APPLICANT: BURNIE, JAMES P
: APPLICANT: MATTHEWS, RUTH C
: TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PILLSBURY, MADISON & SUTRO, LLP
: STREET: 1100 NEW YORK AVENUE, N.W., EAST TOWER, 9TH
: CITY: WASHINGTON
: STATE: DC
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/687,956A
: FILING DATE: 29-JUL-1996

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 19:18:50 ; Search time 22.73 Seconds

(without alignments)
2208.065 Million cell updates/sec

Title: US-09-677-752-4

Sequence: 1 MRPDHNMCCCLCAAILSTTA.....CVLRGQSHSYSLDGLTYRPF 878

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 5716335 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_0401.*
1: /SID52/gcgdata/geneseq/geneseqp/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseqp/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseqp/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseqp/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseqp/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseqp/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseqp/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseqp/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseqp/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseqp/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseqp/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseqp/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseqp/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseqp/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseqp/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseqp/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseqp/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseqp/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseqp/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseqp/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseqp/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseqp/AA2001.DAT.*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4513	99.6	880	21	B13632 C. trachomatis pmp
2	4371	96.4	866	21	B13638 C. trachomatis pmp
3	2006	44.3	438	20	Y30229 Chlamydia trachoma
4	944.5	20.8	928	21	Y90239 Chlamydia antigen
5	940.5	20.7	928	20	W88423 Chlamydia pneumonia
6	930	20.5	927	20	Y35054 Chlamydia pneumonia
7	919.5	20.3	930	20	Y35052 Chlamydia pneumonia
8	912.5	20.1	930	20	W88424 Chlamydia antigen
9	906.5	20.0	930	20	W88420 Chlamydia pneumonia
10	883	19.5	928	20	W88418 Chlamydia pneumonia
11	866	19.1	928	21	Y94327 Chlamydia pneumonia

12	865	19.1	918	21	Y69369	Amino acid sequenc
13	857	18.9	928	20	W88421	Chlamydia pneumonia
14	851	18.8	945	21	Y69368	Amino acid sequenc
15	847	18.7	928	21	Y90237	Chlamydia antigen
16	846.5	18.7	914	20	W88429	Chlamydia pneumonia
17	846	18.7	885	21	Y90238	Mature Chlamydia a
18	846	18.7	949	20	Y35060	Chlamydia pneumonia
19	845	18.6	928	20	W88417	Chlamydia pneumonia
20	845	18.6	928	21	Y90236	Chlamydia antigen
21	843	18.6	918	20	W88422	Chlamydia pneumonia
22	825	18.2	945	20	W88428	Chlamydia pneumonia
23	810.5	17.9	973	21	Y96274	Chlamydia pneumonia
24	794	17.5	182	20	Y37231	Chlamydia pneumonia
25	773	17.1	925	21	Y99843	Chlamydia trachoma
26	773	17.1	936	21	Y99842	Chlamydia pneumonia
27	763	16.8	148	20	Y37230	Chlamydia pneumonia
28	759.5	16.8	1012	20	Y16735	Chlamydia trachoma
29	758.5	16.7	1006	21	B13639	C. trachomatis LGV
30	757.5	16.7	982	21	B13633	C. trachomatis pmp
31	753	16.6	1013	20	Y16737	C. trachomatis pmp
32	749	16.5	1013	20	Y16738	C. trachomatis B s
33	746.5	16.5	922	20	Y34597	Chlamydia pneumonia
34	735	16.2	922	21	W88419	Chlamydia pneumonia
35	735	16.2	922	21	Y95548	Chlamydia pneumonia
36	707.5	15.6	841	21	Y92818	C. pneumoniae CPN1
37	705.5	15.6	841	20	W88420	Chlamydia pneumonia
38	702.5	15.5	1132	20	Y35048	Chlamydia pneumonia
39	664	14.6	643	20	Y35056	Chlamydia pneumonia
40	627.5	13.8	597	20	Y34611	Chlamydia pneumonia
41	576	12.7	671	20	Y35050	Chlamydia pneumonia
42	574.5	12.7	954	21	Y92830	C. pneumoniae CPN1
43	561.5	12.4	788	21	Y92831	C. pneumoniae CPN1
44	558.5	12.3	978	21	Y92834	C. pneumoniae CPN1
45	540.5	11.9	940	20	Y35082	Chlamydia pneumonia

ALIGNMENTS

RESULT	1
ID	B13632 standard; Protein: 880 AA.
XX	XX
AC	H13632:
XX	XX
DT	02-FEB-2001 (first entry)
DE	XX
XX	XX
DE	C. trachomatis pmp1 gene protein.
XX	XX
KW	Chlamydial infection; sexually transmitted disease;
KW	pelvic inflammatory disease; PID; tubal obstruction; infertility;
KW	trachoma; blindness; acute respiratory tract infection;
KW	atherosclerosis; coronary heart disease; antibacterial.
OS	Chlamydia trachomatis.
XX	XX
FT	Key
FT	Misc-difference 336
FT	Location/Qualifiers
XX	/note= "Unspecified amino acid"
PN	MO200034483-A2.
XX	XX
PD	15-JUN-2000.
XX	XX
PF	08-DEC-1999: 99W0-US29012.
XX	XX
PR	08-DEC-1998: 98US-0208277.
PR	08-APR-1999: 99US-0288594.
PR	01-OCT-1999: 99US-0410568.
PR	22-OCT-1999: 99US-0426571.
XX	XX
PA	(CORI-) CORIXA CORP.
XX	XX

XX	26-APR-1999	(first entry)
XX	Chlamydia pneumoniae	surface exposed protein omp10.
DE	Omp10; outer membrane protein 10; surface exposed protein;	
XX	antigen; infection; diagnosis; vaccine; atherosclerosis; asthma.	
KW	Chlamydia pneumoniae.	
XX	WO9858953-A2.	
XX	30-DEC-1998.	
XX	19-JUN-1998;	98WO-DK00266.
PF	23-JUN-1997;	97DK-0000744.
PR	(BIRK/) BIRKELUND S.	
PA	(CHR1/) CHRISTIANSEN G.	
PI	Birkelund S, Christiansen G, Knudsen K, Madsen A;	
PI	Mygind P;	
XX	WPI: 1999-105610/09.	
DR	N-PSDB: X06822.	
PT	Species-specific test for identifying mammals infected with	
PT	Chlamydia pneumoniae - comprises detecting antibodies specific for	
PT	outer membrane proteins of C. pneumoniae or nucleic acids encoding	
PT	these proteins	
XX	Claim 7; Page 60-62; 115pp; English.	
PS	This polypeptide comprises the novel 98.4 kDa surface exposed	
CC	protein Omp10 of the human respiratory pathogen Chlamydia	
CC	pneumoniae. Its amino acid sequence was deduced from DNA (see	
CC	X06822) isolated from a C. pneumoniae expression library. The	
CC	invention provides 12 novel surface exposed proteins, Omp4-Omp15	
CC	(see W88417-28), and nucleic acid sequences encoding them (see	
CC	X06816-27). A new species specific test is claimed that is used	
CC	to identify mammals (including humans) infected with Chlamydia	
CC	pneumoniae. The test comprises detecting antibodies specific for	
CC	Omp4-Omp15 or detecting nucleic acid fragments encoding these outer	
CC	membrane proteins, especially by PCR. The proteins are also used	
CC	in the diagnosis of C. pneumoniae infection in mammals. The	
CC	nucleic acids and proteins can also be used in the immunization of	
CC	mammals, the nucleic acids being particularly useful as DNA	
CC	vaccines for effecting in vivo expression of antigens. The	
CC	vaccines may also prevent atherosclerosis and bronchial asthma,	
CC	which are possibly associated with C. pneumoniae.	

[illegible]

Db	230	asstfsnskafisfnmsvtaatsaagga	lycs--tsapkvyltsdngelnfngtaits	287
Oy	266	GGAIYAKHMYRNRNGPNVSPFINNS----	AKIGAIATIGSGSLSLAGESVLPONNS--	318
Db	268	ggalytdnlvtsagpqlfkmsaidlaap	ygalaialadselsa19gdltfegntlv	347
Oy	319	-QRSSDGLVRNATLYL-EKDAILSLERAND-	ILFDPPIVQESSKESPLPSLAQSVT	375
Db	348	kgasssqrltnslngltlnakivqlasg	ntlylfpdltbht-----aals	396
Oy	376	SPTPTASPLVYIQTASNRSVFSSERLSE	EETK-PDNLTSOLOPIELKSGRLVLDRAV	434
Db	397	dalnngpdlanparygqlvifsgelisc	aeaaadnksrlgqpltaagqalsksgvt	456
Oy	435	LSAPSLSDPPALLIMEXGTSLSKTSDD	KLATLSIPHLSDTEKSVTHAPNLSIQKIFL	494
Db	457	lvakfsfsgpsstllmdgltletadglt	lnlnlvnvdskelkatlkatqas-qvrl	515
Oy	495	SNS----GDENFENVLLSKRONN----	IPULTLSEQ-SHLHPDQNLSS-----	537
Db	516	sgslslvdpbgvnyadv-----swnp	yfscsltltdaddpanihltd--laadpleknp	568
Oy	538	HNGYQSDWFFMSKKDDEGHSLTA--NMT	PKNVPIPEROSTLVANTYSDMAQVOSM	595
Db	569	hngyqgnwalswgedataskkaaltl	twkctgynperrgtlvantlwgstvdvrsiq	628
Oy	596	IMTIAHGAYLLEGTWGSANSLNFVAH	DSKRPIDMHHRSLOGLFGISTHSDHSCFLA	655
Db	629	vatkvrgsgeltgfwegjsnff--hk	dstklmkfmrhsagvvgvattlasdnllta	686
Oy	656	AGOLGKSSDSPTTSTETTSTYATV	QAO-LAT-----SLMK-----ISMAQVYN	698
Db	687	fclqfkgdhdhfnkrraayaasnlh	qplalsspsllrlypgseseqpvlfdqisy	746
Oy	699	ESTHELKTTRYRSKSGFSGFMSVAV	SGVCAPIPV-NGSGLTFSSFIKSLDGSGT	757
Db	747	yskntmctlytqapk-geeswyn	dgcataelass]phtalshegithayrfpkveasy	805
Oy	817	VVLKNAVSDAPMANLDSRA-----	YMRKLTNQRALHRIQTLNANSCVLRGOSHSY	868
Db	866	tailnmswktcgtlnsrqaglg	rfayafs-----pnlvtsnlsmelrgsrsy	918
Oy	869	SLDLGTTRYR	878	
Db	919	nadlgkrlqf	928	
RESULT	6			
YJ35054				
ID	YJ35054	standard: Protein: 927	AA.	
AC	YJ35054:			
XX				
DT	13-SEP-1999	(first entry)		
XX				
DE	Chlamydia pneumoniae surface exposed	polypeptide.		
XX				
KW	Respiratory disease: pneumonia; bronchitis; heart disease; sarcoidosis			
KM	sluhsllats: purulent otitis media; erythema nodosum; pharyngitis;			
XX	vaccline: neutralising epitope.			
OS	Chlamydia pneumoniae.			
XX	W09927105-A2.			
PN	03-JUN-1999.			
PD				
XX	20-NOV-1998:	98MO-IB01890.		
XX				

Query Match	20.3%	Score 919.5;	DB 20;	Length 930;
Best Local Similarity	31.0%;	Pred. No. 1.7e+68;		
Matches 260;	Conservative 150;	Mismatches 343;	Indels 87;	Gaps 26

Qy	98	LSIQNRRPLSF-----TDCSSKSEPSSTIIHKNQGLSLRNNGSSPFK---NNAEISSGGA	149
Qy	119	liltfgsnlsfliaapglvtvasgkscls-----saglnltldnglilifegvnsneannngg	174
Qy	150	ISADAFSLQHNLFLEAFEESSSKGNGCAI-QAOTFSLSRNVSPISFARNRDLNGAICC	208
Db	175	ltakltisgntssltlftsnssakklggalyssaaaslsngntgqvlftmnkxgetggga1--	232
Qy	209	SNLICGNNV---NPLUFTGNSATN-----GCATICISDINTSEKSLSLACQENLFASNS	261
Db	233	-gfeassllqmsllfsgntatdaagkygalc---ektgeiprllisgpkalslftens	288
Qy	262	AKKGAALYAKHNWLYRNGPVSFINN-----SAKIGALAIQSGGSLIIAGESSVLPON	316
Db	289	svtggagicaahgldlaagpqlfsmrcgntaagkgalaladsygsilsaengdclftlg	348
Qy	317	NS-QRTSDQGLVNNAIYLEKDLITLSELRNG-DLTFEDPIVOEESSEKSPSLQASV	374
Db	349	nltltsaptstnralnylgsasakltnlraagqsylyfoplasntgssdvl-----	400
Qy	375	TSPPAPASPLVLIOTSANRSVIRFSERLS-EEEKTRDNLTSLOOPIFLAKSGRLVLKRA	433
Db	401	tlmpdsnspl-----dysgvlivfsgklsadeekaadntfslkqplalaagylalkgnv	456
Qy	434	VLSAPSLSDPPALLMEATSLKTSDD-LKLATLSLPMSIDTEKSVYINAP-----N	486
Db	457	eldvngftqtegsellmqpklkadteaalslklvldlsalegnkvsyleagaekkit	516
Qy	487	LSIQKIFLNSGDNFYEENVELLSKEQNNIPRLTSLSEQSHLDPDGLNS-----HF	539
Db	517	ltspvlfgdsq--nfyeshltlmgafetqplvltfaataasdiyl-dalltspvtqpephy	573
Qy	540	GYGQDMTFSMKDSQEGSLIANNTPKKYVNVHPRROSTLVANLTMYNSDMAQVOSIMINTI	599
Db	574	gqygmweatwedsctaksgmlwtvcltynpnerfssavpdsfwastlflrltqlgmcsq	633
Qy	600	AHGCAYLCTGTSVAVSNLFPAHDSSCKPRIDNMHHRSLGVLFGISTHSLDHSFCILAQOL	659
Db	634	anslyqqrqlwasltanflh-kdksyqln-qafihksyqylvggsaedfsemlfsvafcqj	691
Qy	660	LOKSSDSFTSTETTSYIATVQAO-----IATSLMK-----ISNQCYNES	700
Db	692	fgkddlflvnteshnylaslylqhrflgglpmrpfsgltdmlkdpilflnaqlisysyt	751
Qy	701	HEKLTKTRSPSKKEFGSMHVAASGEVCAISPI-VNNGSGLSSPSFSLQGFSGTOD	759
Db	752	kdmdltirlysy-peggsawctmsgalalagslalyldkeaprlfqqylfplklkdaystrq	810
Qy	760	GFEESSGEIRSFSSAFSPNISLPILGITEFKKSOKTRTYVFLCAYIQDLKRDVESCPLYL	819
Db	811	nlkesgaearatfdgdglvncslpyglrtlekisecedknnfelslalyldvrykrpstrsl	870
Qy	820	LKNASWMDPAMNLDSPAYMRLTNQRLH-RLDTLLNVCGLVIRGOSHSLDGLTTFRF	878
Db	871	mvsqaswtisichnlarqatlasagehltlsphevelsgaayelrtaehlynvdcglrysf	930
RESULT	8		
ID	Y90240	standard; Protein: 930 AA.	
XX	XX	Y90240:	
XX	XX	29-AUG-2000 (first entry)	
DE	XX	Chlamydia antigen CPN100639.	
KW	XX	Chlamydia antigen; diagnosis; infection; community acquired pneumonia	

	KM	therapy upper respiratory tract disease; bronchitis; sinusitis;
	KW	asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma.
	OS	Chlamydia pneumoniae.
	PX	M0200032794-A2.
	PD	08-JUN-2000.
	PE	01-DEC-1999; 99WO-CAM01147.
	PR	01-DEC-1998; 98US-O110339.
	PR	01-DEC-1998; 98US-O110340.
	PR	01-DEC-1998; 98US-O110427.
	PR	01-DEC-1998; 98US-O110428.
	XX	01-DEC-1998; 98US-O110438.
	PA	(CONN-) CONNAUGHT LAB LTD.
	F1	Murdin AD, Oomen RP, Wang J;
	DR	WPI: 2000-412339/35.
	N-	PSDB: A30853, A30854.
	PT	Nucleic acids encoding polypeptide antigens from Chlamydia useful for preventing, diagnosing and treating diseases such as community acquired pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset asthma -
	PS	Claim 16; Fig 7; 17app: English.
	CC	This sequence is a Chlamydia antigen of the invention, designated CC PnJ00639. The nucleic acids (and their complementary sequences) may be used as diagnostic agents for detecting the presence of nucleic acids encoding Chlamydia antigens in samples according to standard methods, and therefore, for diagnosing Chlamydia infections. For example, they may be used as primers and probes for diagnostic polymerase chain reaction (PCR) assays. Antisense sequences may be used to treat infectious expression of the proteins and may be used to produce the protein antigens. The nucleic acids may also be used to produce the recombinant DNA methodologies. The encode according to standard recombinant DNA methodologies. The proteins may then be used as antigens for the production of antibodies (i.e. as vaccines) for preventing infection by Chlamydia. The antibiotics may also be used as diagnostic reagents for detecting infections. Chlamydia is a pathogen implicated in the development of (for example) community acquired pneumonia, upper respiratory tract disease (especially bronchitis and sinusitis), asthmatic bronchitis, adult-onset asthma and acute exacerbations of asthma in adults.
	SO	Sequence 930 AA:
	OY	Query Match 20.1%; Score 912.5; DB 21; Length 930; Best Local Similarity 30.8%; Pred. No. 6, Se 68; Matches 259; Conservative 149; Mismatches 345; Indels 87; Gaps 26 98 LSTONRFFLSF-----TDCSSKESPPSIHOKNGQLSRNNGSMFCR--NHAECSGGA 149 : : : : : : : : : Dd 119 lltfglnslfiapgtltaasgkstls----sgalnldnglltsgvnseannnga 174 ISADAPSLQNNVLETFATFNSSSKONGCAI-QAOFSLSNRVSPITPARNFADLNGAICC 208 OY 150 ISADAPSLQNNVLETFATFNSSSKONGCAI-QAOFSLSNRVSPITPARNFADLNGAICC 208 Dd 175 ittktltsigntslftltnsakklgaivsaaaisigtqjvlmnnkecgggal-- 232 OY 209 SNLICSGNV---NFLFTGSATIN-----GGAICISDLNTSEKGSLALACNOETLRASNS 261 Dd 233 -gfcaassatnsilfsgnlataadagkgaiyc--ektgetptlitgsnkxsltfaas 288 OY 262 AKEGGAIYA KMHLRRNGRPSPFINN-----SAKIGGANIAIOSGGSLSIJAGEGVLFON 316 Dd 289 svitggagaicalmgdlisaagrpllsnmrcgnhtaagkgataladagsislssangddiflg 348 OY 317 NS-QRTSOGGLVRNAVILEKDAAIISSLERNRG-DILFPDIYDESSSKESPPLDSLIQAIV 374

PR 20-AUG-1998; 98US-0097195.
 PR 20-AUG-1998; 98US-0097196.
 PR 20-AUG-1998; 98US-0097197.
 PR 27-AUG-1998; 98US-0097191.
 PR 17-AUG-1999; 99US-0376770.
 XX (CONN-) CONNAUGHT LAB LTD.
 PA Murdin AD, Oomen RP.
 PI WPI: 2000-224703/19.
 DR N-PSDB: 261509.
 XX Novel antigens and corresponding DNA molecules that can be used to
 PT prevent, treat and diagnose disease caused by Chlamydia infection in
 PT mammals, especially humans.
 PS Claim 19; Fig 15-E: 201pp; English.
 CC Y69362-69 represent Chlamydia pneumoniae polypeptides. The polypeptides
 CC are present in the bacterial membrane structure, in the external
 CC vicinity of the membrane structure, in the inclusion membrane
 CC structure, in the external vicinity of the inclusion membrane
 CC and in the cytoplasm of the infected cell. The polypeptides may be
 CC used to prevent, treat and detect the presence of Chlamydia infection
 CC and/or the presence of Chlamydia in a sample. The polypeptides may
 CC also be used to induce an immune response in a mammal. The vaccine
 CC vector comprising the polypeptides is used to induce an immune
 CC response in a mammal. Antibodies directed against the polypeptides
 CC may also be used therapeutically to treat and/or prevent a Chlamydia
 CC infection.
 CC
 SQ Sequence 918 AA:
 Query Match 19.1%; Score 865; DB 21; Length 918;
 Best Local Similarity 31.2%; Pred. No. 6.6e-64;
 Matches 253; Conservative 138; Mismatches 326; Indels 94; Gaps 29;
 OY 120 IHHKNGQLSLRNNNSFCRNHAEHSGGSAISADAFSLQHNHYLTFAEENSKNGCAIQ 179
 DB 150 lvesignldlnenasse-----nggainktklstlgtstrfvalignsssgg99a1y 201
 OY 180 AQRTS-USRNPSPISFARNRADLNGAICCC-SNLCGNNPPLPPTG-NSATNGCAICCI 236
 DB 202 asgdsvisenagllstignsatstsgaalsagunlvismnq-llfdgckatnggaic- 259
 OY 237 SDLMTSEKSLSLACNOETLFASSNAKEGAIYAKHNVLR-NGPVSFINNSA---KI 291
 DB 260 nkagandpplilistigneshlflmtagnsga1yckk1vlsgrgylfsmnkaanatkp 319
 OY 292 GGAIAIGSGSLISLAGEGSLVLPNNNSORT--SDOGLVRNAIYLEKDALISLEARNND- 348
 DB 320 ggaiallidsgeisladaignllfegntstgspasvtrnaidlsnakflinlratrnk 379
 OY 349 ILFPPIYOESSKSPPLSLQASVTSPPTATSPPLVLOTGANS-----VIF 397
 DB 380 vifgpl-----tssgatklslnxadagsgnllygylvf 414
 OY 398 SEERLSSEEE-KTPDNLTLOOPIELKSGRLVLDRAVLASPSLQDPQALLIMEAGTSL 456
 DB 415 sgeklseeelkphnkhkftgave1aaga1vldgvtvntlltvegskvmdgyltf 474
 OY 457 KTTSD-LKLTATISPLHSID-TEKSV---THAPNLSIOKIFLSNGDGFENVELLSK 511
 DB 475 eesaeagvltng1a1ndsdgtnkalikaaskaakdvalagimlvdaagnyehnm1--s 532
 OY 512 EONNPIPLTLTSMK--QSHMLPDP--NLSHFYGGDMTFEWSKQDDEGSLIA--NMTP 564
 DB 533 gqgvfplle1sagqtm1tdp1p1lntc1nygqgnm1lvvddatataktkaatlwtck 592
 OY 565 KNYVPEROSTLVANTLNNTYSDMOAVQSMITIAHGAYLFGTWSAVSNLFYVHDS 624

DB 593 tgyknpnpergpllvpsn1vsgfvdvrs1qslmdrstss1ssstnlvsg1ad-f1hbedqk 651
 OY 625 GRPIDNMHRRSGYLFEGISTHSLDHSFCLAGCOLLGRKSSDSFIISTETSYIATVQAO- 683
 DB 652 gmq-rsyrlhsagyal99gfflasenfinfalcqilgydkdhlvaknhbhvagaagayrh 710
 OY 684 --LATS1MKI-----SAQACYNES1HELTCKYRSFKEGFGSMHSAVANGEVYC 729
 DB 711 lgeskl1ak1lsgnsd1p1vfnar1ayghldnm1tk1tygspxk-gawgnadafg1ecg 769
 OY 730 ASIP1VSNQ-SGLFSSFS1FKLQGFSGTQDGFEESSGGEIRSFSA5FRN1SLPITFE 788
 DB 770 galpvasgr1swdch1p1nlem1yahndf1keng1eg1rsf1sed1f1n1avp1y1kte 829
 OY 789 KKSQKTRTY1YELGAY1ODLKRDVEGSPVVLKNVSMDBAN1DSRAVMEFLTQORAL 848
 DB 830 kfcdks-tydlst-ayvpdlr1ndpgc1t1lmsvgdsw1stc1st1sarqal1lvag1nha1 887
 OY 849 -HRLQTL1ANVSCV1LKGQSHS1SLD1GTYRF 878
 DB 888 asnfefvsg1evelfrgs1rsyald1g9r1gf 918
 RESULT 13
 ID W88421 standard: Protein: 928 AA.
 W88421:
 AC W88421:
 XX 26-APR-1999 (first entry)
 DT
 XX Chlamydia pneumoniae surface exposed protein Omp8.
 DE
 XX Omp8; outer membrane protein 8; surface exposed protein; antigen;
 KM infection; diagnosis; vaccine; atherosclerosis; asthna.
 KW
 XX Chlamydia pneumoniae.
 OS W09858953-A2.
 PN
 PD 30-DEC-1998.
 XX
 PF 19-JUN-1998; 98WO-DK00266.
 XX
 PR 23-JUN-1997; 97DK-0000744.
 XX
 PA (BIRK/) BIRKELUND S.
 PA (CHR1/) CHRISTIANSEN G.
 XX Birkelund S, Christiansen G, Knudsen K, Madsen A;
 PI Mygind P;
 XX
 DR WPI: 1999-105610/09.
 DB N-PSDB: X06820.
 XX
 PT Species-specific test for identifying mammals infected with
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
 PT these proteins
 PS Claim 7; Page 53-55; 115pp; English.
 PS This polyepitide comprises the novel 90.0 kDa surface exposed
 CC protein Omp8 of the human respiratory pathogen Chlamydia
 CC pneumoniae. Its amino acid sequence was deduced from DNA (see
 CC X06820) isolated from a C. pneumoniae expression library. The
 CC invention provides 12 novel surface exposed proteins, Omp4-Omp15
 CC (see W88417-28), and nucleic acid sequences encoding them (see
 CC X06816-27). A new species specific test is claimed that is used
 CC to identify mammals (including humans) infected with Chlamydia
 CC pneumoniae. The test comprises detecting antibodies specific for
 CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
 CC membrane proteins, especially by PCR. The proteins are also used

```

OY 128 LSLRNNGSMSPCRNHAEGSGAIS-ADAFSLQHNLYFTAFENSSKNGCAIO---AQTF 183
DB 175 lfgiknsaifgntskkqgastlqgltemijtklknkavtsggalidgaastf 234
OY 184 SLRNVSPISFARNRAD---LNGCAICCS-NLIGSGNVNPLEFTGSAINGAICISDL 239
DB 235 tanhe---lfsqnktsngaangaincsqdlftdntslilqenstmgdgaic----- 286
OY 240 NTSEKGSLSLACQETLFPASNAKEKGAIYAHM-VLRINGEVSFINN---SARIGCA 294
DB 287 ---stglstltsqdslnvigtsgkqgaisaaslkllgqgagallsnvthaltprga 343
OY 295 IAOOGSGSLILAGESVLPONNSQRTSDGLV--RNAIYLEKDAIISLEARNCD-ILF 351
DB 344 lftngqsglftcggqdlvteqavtltapnatktrvlnheslaktglaaaggnalyf 403
OY 352 FDPVQESSKESPLPSSLOAVTSPPATASPLVIQTSANR---SVIFSSERLSEEEK 407
DB 404 ydpi-----ltndtgaadnlfrinevsanqklsqsvifsgerlstaee 445
OY 408 TPDLNLSOLOQPIELKSGRLVLDRAVLASPSLQDPOALLIMAGTSKTS--DLKAT 466
DB 446 laeoltarlnqpvltvsgslvltqglvltqglstqdestlildqtslkastediavl 505
OY 467 LSLPLHSLDTEKSVTIAH---PMLSIQKIFLSNSGDENFENELUSKEONNIPLAT-- 520
DB 506 lsnadliygknpnlvsaankliltglalvnaadgaftymhltqdsqstvklsqg 565
OY 521 ---LSKEOSHLLHPGONLSHFQYOGDMTFSMKDSDEGSLIAN--WPKKNVPIPER 573
DB 566 aggtlftqdasqkqlvavapsrphlygqghmvqvlpgtqpsqanleavrtgylpmp 625
OY 574 QSLVANTLMTYSDMOAVQSMINTIAHGAVLCTGWSANSLFYAIDSSGKPRIDMHH 633
DB 626 qgslypsmlwgsfvdqraigeimwnsqllcqezyvwgaglan-llhcdkline--hgyr 682
OY 634 RSLQYLPQISTHSLDDHFCFLAAGQLGKSSDSPTSTETTSYIATV----- 680
DB 683 sgvgylvgvgtahsfadinafcqlfstrdkdyvsknhglsysgvltledlstrpg 742
OY 681 ---QAOLATSLMKISAOCYNESIHETKTRYSFSEKGEFSMHSVAVSGVCASIP 734
DB 743 fytssaseacnqvvtlimglqysnmndmcklycty-pegagswandvfgjefatly 801
OY 735 VSNSGGLFSFISFKLQCFSTOGFESESEIRSFASFRNLSLPRGIFFEKKSQKT 794
DB 802 ypsnftlfdyvspflrlqctyahgedfketgevrhftsgdldfnlavprlqvktfstr 861
OY 795 RPYVFFLAYIOD-LKRDVEGPPVLLKNAVSDAPMANLDSRAVFRLTNORALH-RLQ 852
DB 862 rgsyeltlayvdpvltkdpks--tatlagaatwstghnlsrtgqlrltgnhcllnpgle 919
OY 853 TLLNVSCVLRGQSHSYSLDLGTYRF 878
DB 920 vftshgaletgrsrmynlnlgkyrf 945

```

```

RESULT 15
ID Y90237 standard; Protein: 928 AA.
XX Y90237;
XX 29-AUG-2000 (first entry)
XX Chlamydia antigen CPN100635.
XX Chlamydia antigen CPN100635.
XX Chlamydia antigen: diagnosis: infection; community acquired pneumonia;
XX therapy: upper respiratory tract disease; bronchitis; sinusitis;
XX asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma.
XX Chlamydia pneumoniae.
XX

```

```

FH Key Location/Qualifiers
FT Peptide 1..43
FT Protein /note="signal peptide"
FT Protein 44..928
FT Protein /note="mature CPN100635"
MO2000032794-A2.
PD 08-JUN-2000.
PE 01-DEC-1999: 99MO-CA01147.
XX
XX 01-DEC-1998: 98US-0110339.
XX 01-DEC-1998: 98US-0110340.
XX 01-DEC-1998: 98US-0110427.
XX 01-DEC-1998: 98US-0110428.
XX 01-DEC-1998: 98US-0110438.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Mordán AD, Oomen RP, Wang J;
XX WPI: 2000-412339/35.
XX N-PSDB: A30849; A30850.
XX
XX Nucleic acids encoding polypeptide antigens from Chlamydia useful for
XX preventing, diagnosing and treating diseases such as community acquired
XX pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset
XX asthma.
XX
XX Claim 16: Fig 3: 174pp: English.
XX
XX This sequence is a Chlamydia antigen of the invention, designated
XX CPN100635. The nucleic acids (and their complementary sequences) may be
XX used as diagnostic agents for detecting the presence of nucleic acids
XX encoding Chlamydia antigens in samples according to standard methods,
XX and therefore, for diagnosing Chlamydia infections. For example, they may
XX be used as primers and probes for diagnostic polymerase chain reaction
XX (PCR) assays. Antisense sequences may be used to treat infections. The
XX expression of the proteins and may be used to produce the protein antigens they
XX nucleic acids may also be used to standard recombinant DNA methodologies. The
XX encode according to standard recombinant DNA methodologies. The
XX proteins may then be used as antigens for the production of antibodies
XX (i.e. as vaccines) for preventing infection by Chlamydia. The
XX antibodies may also be used as diagnostic reagents for detecting
XX infections. Chlamydia is a pathogen implicated in the development of
XX (for example) community acquired pneumonia, upper respiratory tract
XX disease (especially bronchitis and sinusitis, asthmatic bronchitis,
XX adult-onset asthma and acute exacerbations of asthma in adults.
XX
XX Sequence 928 AA:

```

```

Query Match 18.7%: Score 847; DB 21; Length 928;
Best Local Similarity 30.1%; Pred No. 2.3e-62;
Matches 284; Conservative 141; Mismatches 409; Indels 108; Gaps 31;

```

```

OY 12 CAIISSTAVLFGODPL--GETALLTKPNHNVV---TFEEDCTNESL--FPALCAHAS 63
DB 20 cstvfataenipsgsfsgstntgtytknttgdtyltgltlqngydsaalckgct 79
OY 64 QD--DPIYVLGNSYCPVSKLHITDPEALFKKGLDISIONFFELSTDCSSKKE--SSPS 119
DB 80 sdtesistagkxys--lsflnkssaeq-----aalsvtdknlsitgfsllflaaps 132
OY 120 Ii-----HOKNGOLSLRNNGSMSPCRNHAEGSGCAISADAFSLQHNLYFTAFEN 170
DB 133 svltlpsgkavxcgaldtdmgtllfkqyceenggaistknlskntsgistegnk 192
OY 171 SKG---NGCAIOAO-TFSLRNVSPISFARNRADLNGAI-CCSNLIGSGNVNPLEFTGN 225
DB 193 ssatgkkggataicgtvdlcnntaprlfsmiaaaggaistngcttgnts-lyfsen 251

```

```

OY 505 NVELLSKEQN-----NIPILITLSKEQSHLHPDGNLSHFYOGDWTFSWKD-----SDEGH 556
DB 531 nhd-lygktqdfsfvgjsaigtattdvpavrpvaepthgygvtgmwtvddtaetpkx 589
OY 557 SLIAWTPKNVYHPHEROSTLVANTIMNTYSDMAQVQSMINTIAHGAYLFTQWGSANVN 616
DB 590 fatlwnttgyipmgrrgpylpnslwgsfsgdqlqglersalltlcsdrgfwaagyan 649
OY 617 LFYADHSSGKPIDNMHNSLGLFGLSTHSLDHSCLAGQLCKSSDSFITSSTTTSY 676
DB 650 -fldkdkkgek-rykyrhksgyaigaaqtcsehlisfaicqifgsdkxfivaknhtdy 707
OY 677 IATVQAQLATSL-----MKISAQACYNESIHBLKTKYRSFSEKGG 717
DB 708 agafyqhhtecsgfgyclldklpgawshkplvegqlayshvsnldkkytaty-pevxyg 766
OY 718 SMHSVAVSGEVCASIPIVNGSGLFSSGFIFSKLQGFSGTODGFERSGGEIRSFASSPR 777
DB 767 swgnafmmiigasshsypelyhctfityapyiklnltlytqdsfsekgtgtsfddsnlf 826
OY 778 NISLPIGITEFEKKSQKTRTYVFLGAYIDLKRDVESGPVYLKNAVSDAPMANIDSR 837
DB 827 nlsipgvftekfscdndfsydlitlsyvpdlirndpckctalvisgawetlyannlartga 886
OY 838 YMER 841
DB 887 lqyr 890

RESULT 17
Y90238
Y90238 standard; Protein: 885 AA.
Y90238;
XX
AC Y90238;
XX
DT 29-AUG-2000 (first entry)
XX
DE Mature Chlamydia antigen CPN100635.
XX
KW Chlamydia antigen; diagnosis; infection; community acquired pneumonia;
KM therapy; upper respiratory tract disease; bronchitis; sinusitis;
KM asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma.
XX
OS Chlamydia pneumoniae.
XX
PN W0200032794-A2.
XX
PD 08-JUN-2000.
XX
PE 01-DEC-1999; 99WO-CA01147.
XX
PR 01-DEC-1998; 98US-0110339.
PR 01-DEC-1998; 98US-0110340.
PR 01-DEC-1998; 98US-0110427.
PR 01-DEC-1998; 98US-0110428.
PR 01-DEC-1998; 98US-0110438.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Murdin AD, Oomen RP, Wang J;
XX
DR WPI: 2000-412339/35.
DR N-PSDB: A30849, A30850.
XX
PT Nucleic acids encoding polypeptide antigens from Chlamydia useful for
PT preventing, diagnosing and treating diseases such as community acquired
PT pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset
PT asthma.
XX
PS Claim 16; Fig 3; 174pp; English.
XX
CC This sequence is a Chlamydia antigen of the invention, designated
CC CPN100635. The nucleic acids (and their complementary sequences) may be

```

```

CC used as diagnostic agents for detecting the presence of nucleic acids
CC encoding Chlamydia antigens in samples according to standard methods,
CC and therefore, for diagnosing Chlamydia infections. For example, they may
CC be used as primers and probes for diagnostic polymerase chain reaction
CC (PCR) assays. Antisense sequences may be used to down regulate
CC expression of the proteins and may be used to treat infections. The
CC nucleic acids may also be used to produce the protein antigens they
CC encode according to standard recombinant DNA methodologies. The
CC proteins may then be used as antigens for the production of antibodies
CC (i.e. as vaccines) for preventing infection by Chlamydia. The
CC antibodies may also be used as diagnostic reagents for detecting
CC infections. Chlamydia is a pathogen implicated in the development of
CC (for example) community acquired pneumonia, upper respiratory tract
CC disease (especially bronchitis and sinusitis, asthmatic bronchitis,
CC adult-onset asthma and acute exacerbations of asthma in adults.
XX

SQ Sequence 885 AA:

Query Match 18.7%; Score 846; DB 21; Length 885;
Best Local Similarity 30.5%; Pred. No. 2,5662;
Matches 279; Conservative 135; Mismatches 382; Indels 120; Gaps 30;

OY 28 LGETALLTKPNINIVCTFFEDCTMESLFPALCAHSDPDLYLVGNISYCFWFKLHTDP 87
DB 25 lqdsaaalkg-----c--fsdlt-----eslsfagkyg--lsflnkss 60
OY 88 KEALFKKKGDSIONFFRLFTDCSKE--SSPSII-----HOKNQOLSRNNGSM 136
DB 61 aeg-----aalsvtdknslstgtslflaapsvltlpsgkavvcgdlfdmngtl 115
OY 137 SFCRNHAEGSGAISADAFSLQHNLYLTAPEENSSK-----NGAIOQO-TPSLSRNVSPI 192
DB 116 lfkgdyceenggaistknlslknsstsfegnkssatgkkgacatgvtldtmnapt 175
OY 193 SFARNRADLNGAL-CCSNLCISGNVNPFLTFTGNSAT-----NGAICISDLNTEKGS 247
DB 176 lfsmiaaagaalnsitnctigtnts-lvisensvatlaagngal-----sgadv 226
OY 248 SLACNOETLFAFNSAKEKGAIVAKHMYLRVNGHVS--FINN-----SAKIGATAIQSG 300
DB 227 llsqgsvtfsngqavangaalylakklilaagggggnpfsnmlyqgtlaagngasllaa 286
OY 301 GSLSLAGEGVLFQNNNS-ORTSDQGLVRNALYLEKDALLSLEARNG-DLFFDPPIVOE 358
DB 287 gecslfseagchylngnalvatlptqlkrnsldtstgkdhelirajsglsifdyptlan 346
OY 359 SSSKESPLPSLQASVTSPTPATASPLVIQTSANRSVIFSSPRLSEEE-KTPDNLTSOLO 417
DB 347 laadstdtlnlnkadagn-----stlqysglvfygsklsedekavadtlnstlkl 395
OY 418 OPTELKSGRLVKDRAVYAPSLSDOPALIMEAGTSILKTES-DUKLATISIPLHSDT 476
DB 396 gpvtlaagdlvklrgvltldtkgftqtagssvlimdgutlkatleevltlgslpdslyge 455
OY 477 EKSVTIIHAP-----NLSIOKIFLNSGDENFENVVELLSKEQN-----NIPILITLSKEQS 526
DB 456 gkkvylaasaaknvalsgpdlldngq--nayenhd-lygktqdfsfvgjsaigtattd 512
OY 527 HHLHPDGNLSHFYOGDWTFSWKD-----SDEGHSLIAWTPKNVYHPHEROSTLVANTL 582
DB 513 vpavrpvaepthgygvtgmwtvddtaetpkktktalwnttgyipnpergplvpnsl 572
OY 583 WNTYSDMAQVQSMINTIAHGAYLFTQWGSANVNLFYADHSSGKPIDNMHNSLGLFGL 642
DB 573 wgsfsgdqlqglersalltlcsdrgfwaagyan-fldkdkkgek-rykyrhksgyaig 630
OY 643 STRSLDHSFCLAGQLCKSSDSFITSSTTTSYIATVQAQLATSL----- 688
DB 631 aagtcsehlisfaicqifgsdkxfivaknhtdyagaifyqhhtecsgfgyclldklpgs 690
OY 689 -----MKISAQACYNESIHBLKTKYRSFSEKGGSMHSVAVSGEVCASIPIVNGSGLFS 743

```


CC	which are possibly associated with C. pneumoniae.									
XX	Sequence	918 AA;								
SQ	Query Match	18.6%; Score 843;	DB 20;	Length 918;						
	Best Local Similarity	31.2%; Pred. NO. 4.8e-62;								
	Matches	249;	Conservative 137;	Mismatches 325;	Indels 86;	Gaps 28;				
QY	134	GSMFCRHAEGSGAISADAFSLQHNVLFTAFENSKNGGAIQAOTFS--LSRNVSP	192							
DB	156	gnldqnenassengaintklsitgrfvaflgnssqgggaiyasgdsvisenagil	215							
QY	193	SFARNRDLNGGAICC--SNLICSGNVNPLFTFG--NSATNGAICISDLNTEKGSLSLA	250							
DB	216	sfgnssattsagaisaegnvlisnnqn--ifidgckattnggaide--nkaganpdpillts	273							
QY	251	CNQTFLFASNAKEKGGAIAKHMVLRV--NGPVSPFINNSA----KICGAIAIQSGSLSI	305							
DB	274	gneslhnntagnsggaiytkkvlssgrgvgifsnkkaanatkpggaialldseis	333							
QY	306	LAGEGSLVFONNRSOT--SDQGLVNAIYLEKDAIILSLSEARNGD--ILFEDPIVOESSK	362							
DB	334	sadlnilfegnttstspasvtrnaidlasnkflnratrgnkvlfydpl-----	386							
QY	363	ESPLPSSLOQASVTPTATASPLVLOTSAHRS-----VIFSPRSEEP--KTPD	410							
DB	387	-----tssgatklsinkadagsgntyegylvfsgklseseeeklkp	428							
QY	411	NITSLOQPILKGRVLKDRAVLSAPLSODPQALLIMEAGTSLKTSDD--LKLATLSI	469							
DB	429	nikstftgavalaagvlkdgvtvvanitqvegskvmdggttfeasaegvtnglai	488							
QY	470	PLHSLD--TEKSV--TIHAPNISIOKIFUSGDNFENVELLSKEQNIPLLTLSK--	524							
DB	489	nldslgtkalkaikaasakvaigpmlvdaqngnyehnl--sqgqvflieisaag	546							
QY	525	-OSHLHLPDQ---NLSSHFYGOQDWTFSMKDSDEGHSLIA--NWTPKNYVPHPEROSTLV	578							
DB	547	tmttdipdptlntttnhygggtglivwvdataaktknatllwtktg/kpnperrgplv	606							
QY	579	ANTLWNTYSDMQAVQMINTIAHGGAYLFGTWCASVSNLIFYAHDSSCKPIDNWHHSIGY	638							
DB	607	pnaIwgsfvdvrsiqslmdrtsslsstnIwvsgiad-flhedqkng-rysrhsag	664							
QY	639	LFGISHTSLDDHSCFAAGCOLCKSSDSFTTSTTTSYIATVQAO---LATSLMKI----	691							
DB	665	alggrftasenfnfafcqlfygdhvaknhtvgegamsyrhlgesktlakilsgn	724							
QY	692	-----SAQACYNESIHELKTKYRSFSGEFCSMHSAVSGEVCASIPVNSG--SGLF	742							
DB	725	sdsIpfvfnarfayghtdnmttkytsypvk--gswgndafglecggaipvvasgrsw	783							
QY	743	SPSIFSKLQSGCTQCPRESSGCEIRSPSASSFRNLISLIGITFEKSKTKRTYYFLG	802							
DB	784	dthcpfnlemyahqndfkengtegrsfqsdelfnlavpvvgikfekfsdks-tydlisi-	841							
QY	803	AYIOTDLKRDVESGPVLLKNVSDNAPMANLDSRAYMFLTNQRAL--HRLQTLNNSCVL	861							
DB	842	ayvpdvirndpgcttllmvsgdsvstcgtslsrqallivregnhhafasnfefvsqfevel	901							
QY	862	RGQSHYSLSLDLGTYYRF	878							
DB	902	rgsrsyaidlggrgfg	918							
RESULT	22									
W88428										
ID	W88428	standard; Protein;	945 AA.							
AC	W88428:									
XX										
DT	26-APR-1999	(first entry)								

XX	Chlamydia pneumoniae surface exposed protein Omp15.									
DE	Omp15: outer membrane protein 15; surface exposed protein;									
XX	antigen; infection; diagnosis; vaccine; atherosclerosis; asthma.									
KW	Chlamydia pneumoniae.									
OS	WO9858953-A2.									
XX	30-DEC-1998.									
PD	19-JUN-1998;	98WO-DK00266.								
XX	23-JUN-1997;	97DK-0000744.								
XX	(BIRK/) BIRKELUND S.									
PA	(CHRI/) CHRISTIANSEN G.									
XX	Birkelund S, Christiansen G, Knudsen K, Madsen A;									
P1	Mygind P;									
XX	WPI; 1999-105610/09.									
DR	N-PSDB; X06827.									
XX	Species-specific test for identifying mammals infected with									
PT	Chlamydia pneumoniae - comprises detecting antibodies specific for									
PT	outer membrane proteins of C. pneumoniae or nucleic acids encoding									
PT	these proteins									
XX	Claim 7; Page 71-73; 115pp; English.									
XX	This polypeptide comprises the novel surface exposed protein Omp15									
CC	of the human respiratory pathogen Chlamydia pneumoniae. Its amino									
CC	acid sequence was deduced from DNA (see X06827) isolated from a C.									
CC	pneumoniae expression library. The invention provides 12 novel									
CC	surface exposed proteins, Omp4-Omp15 (see W88417-28), and nucleic									
CC	acid sequences encoding them (see X06816-27). A new species									
CC	specific test is claimed that is used to identify mammals									
CC	(including humans) infected with Chlamydia pneumoniae. The test									
CC	comprises detecting antibodies specific for Omp4-Omp15 or detecting									
CC	nucleic acid fragments encoding these outer membrane proteins,									
CC	especially by PCR. The proteins are also used in the diagnosis of									
CC	C. pneumoniae infection in mammals. The nucleic acids and proteins									
CC	can also be used in the immunization of mammals, the nucleic acids									
CC	being particularly useful as DNA vaccines for effecting in vivo									
CC	expression of antigens. The vaccines may also prevent									
CC	atherosclerosis and bronchial asthma, which are possibly associated									
XX	with C. pneumoniae.									
SQ	Sequence	945 AA;								
	Query Match	18.2%; Score 825;	DB 20;	Length 945;						
	Best Local Similarity	28.2%; Pred. NO. 1.7e-60;								
	Matches	244;	Conservative 153;	Mismatches 371;	Indels 98;	Gaps 25;				
QY	74	SYCMFVSKLHITDPKEALFKKGDGDL-----STONFRLFTDCSSKESPSIIHOKNGO	127							
DB	117	sftdfslvitespkpsavstgkgsilvssgavqlqdlntlvitsasvedgvgi--kgnsc	174							
QY	128	LSLRNCSMSFCRHAEGSGGAIS--ADAFSLQHNVLFTAFENSKNGGAIQ---AOTF	183							
DB	175	liqgiknsaifgqntskkkggaisttqglitienlgtlkfneknkavissggaiddgaastf	234							
QY	184	SLSRNVSPISFARNRAD---LNGGAICCS--NLICSGNVNPLFTGNSATNGAICCSIDL	239							
DB	235	tanhe---lifspkntsgnaangaincsgdlftdntslilqenstmqdggaic-----	286							
QY	240	NTSEKGSLSLACNOETLTFASNAKEKGGAIAKHM-VLRYNGPVSPFINN-----SAKIGA	294							
DB	287	---stgttisigdsdinvgitsgdkggaisaasikilggggalifsnvntthatplgga	343							

OY 655 AAGOLLGKSSDFITSTETTSYATVQAOATSLMKIS-----AAQYNESIHELTK 707
DB 742 gfgqltkskdylvgghghnvfatvynithkslfgsffggtsrvtsrnekvks 801
OY 708 YRSFSGEGFSGHSHVAVSGEVCASIPVNGSGL-FSSFSIFSFKLQFGSGTQDGFESSG 766
DB 802 ytklpk-grcswnncwigelegnlpitlssrllnkqilpfvkaevayathgqgentp 860
OY 767 EIRSFASFSRNISLPITGTFEKKSKQTRTYVFLGAYIQDLKRDVESGPVLLKNAVSW 826
DB 861 egrifghhllnvavpvgvrfgknsnrpdytilvayapdvyrhnpdcdttlpingatw 920
OY 827 DAPMANLDSRAYMRLTNORALHR-LQTLNVSCVLRGSHSYSLDITGTYRF 878
DB 921 tsignlstrllvqasshslsvndvleifghcgcdirtsrkylvldigsklrf 973

RESULT 24
Y37231
ID Y37231 standard; Protein; 182 AA.
AC Y37231;
XX
XX
DT 07-OCT-1999 (first entry)
XX
DE Chlamydia trachomatis cellular envelope protein.
XX
XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KW Bartholinitis; pneumopathy; venereal lymphogranulomatosis.
XX
OS Chlamydia trachomatis.
XX
PN WO9928475-A2.
XX
PD 10-JUN-1999.
XX
XX 27-NOV-1998; 98WO-IB01939.
XX
PR 04-NOV-1998; 98US-0107077.
PR 28-NOV-1997; 97FR-0015041.
PR 17-DEC-1997; 97FR-0016034.
XX
PA (GEST) GENSET.
XX
PI Griffais R;
XX
DR WPI; 1999-371125/31.
XX
PT Genome sequence of Chlamydia trachomatis
XX
PS Disclosure: Page 992; 1755pp; English.
XX
CC Y36754-Y37949 are encoded by open reading frames (ORFs) of the genome of
CC Chlamydia trachomatis (see 201425). The polypeptides can be used as
CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC can also be used to control growth of the microorganism. Chlamydia
CC trachomatis is responsible for a large number of diseases, e.g. eye
CC diseases such as conventional trachoma, nonendemic trachoma,
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
CC Bartholinitis, Bartholinitis; pneumopathy in breast feeding infants;
CC and venereal lymphogranulomatosis. The polypeptides of the invention
CC may be of use in treating these diseases.
XX
SQ Sequence 182 AA;

Query Match 17.5%; Score 794; DB 20; Length 182;
Best Local Similarity 97.5%; Pred. No. 4.1e-59;
Matches 157; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 705 KTKYRSFSGEGFSGHSHVAVSGEVCASIPVNGSGLFSSFSIFSFKLQFGSGTQDGFESS 764
DB 1 knidgsfsgksgwhsvavsgvscasipvngsglfsfsfslqgsgtqdgfess 60
OY 765 SGEIRSFASFSRNISLPITGTFEKKSKQTRTYVFLGAYIQDLKRDVESGPVLLKNAV 824
DB 61 sgeirsfasfsrnislplgltfekkksqtrtyyflgavlyqdlkrdvesgpvllknv 120
OY 825 SWDAPMANLDSRAYMRLTNORALHRLQTLNVSCVLRGOS 865
DB 121 swdapmanldrsymrltnqralhrlqtltnvscvlrgds 161

RESULT 25
Y99843
ID Y99843 standard; Protein; 925 AA.
XX Y99843;
XX
XX 15-SEP-2000 (first entry)
XX
DE Chlamydia pneumoniae processed 98 kDa outer membrane protein CPN100640.
XX
KW Chlamydia; 98 kDa outer membrane protein; antigen; immunogen; infection;
KW vaccine; antibacterial; community acquired pneumonia; bronchitis;
KW sinusitis; acute respiratory disease; upper respiratory tract disease;
KW asthma; atherosclerosis.
XX
OS Chlamydia pneumoniae.
XX
PN W0200032784-A1.
XX
PD 08-JUN-2000.
XX
XX 01-DEC-1999; 99WO-CA01148.
XX
PR 01-DEC-1998; 98US-0110439.
PR 03-MAY-1999; 99US-0132272.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Murdin AD, Oomen RP, Wang J, Dunn P;
XX
DR WPI; 2000-412330/35.
DR N-PSDB; A48838, A48839.
XX
PT New polynucleotide encoding the Chlamydia 98 kilodalton outer membrane
PT protein, useful for preventing or treating Chlamydia infection -
XX
XX Claim 16; Fig 1; 98pp; English.

The present sequence is the mature, processed form of CPN100640, the
98 kDa outer membrane protein from Chlamydia pneumoniae. Chlamydia
pneumoniae is a common cause of community acquired pneumonia and upper
respiratory tract symptoms and diseases, including bronchitis and
sinusitis. It also has an association with atherosclerosis and asthma.
The 98 kDa outer membrane protein is a C. pneumoniae-specific antigen
which can confer immune protection against chlamydial infection. The
nucleotide sequence encoding the protein or the protein itself may be
administered as a vaccine to prevent or treat infection and they may
also be used to diagnose infection. The gene encoding CPN100640 was
amplified from Chlamydia pneumoniae genomic DNA by PCR.

Query Match 17.1%; Score 773; DB 21; Length 925;
Best Local Similarity 28.0%; Pred. No. 4.1e-56;
Matches 236; Conservative 143; Mismatches 378; Indels 86; Gaps 22;

OY 97 DLSTONFRFLSFDTCSKESSPS--IIHQKNGQLSLRNGSMSCFRNHAEGSGGAISADA 154
DB 108 nllfnfdrslscpslllspptgqcalksvgnlsitgnsqilftqnfssdnggvintkn 167

OM of: US-09-677-752-4 to: N_Geneseq_0401.* out_format : pfs
Date: May 6, 2001 8:37 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+P2n.model -DEV=xlp
-O=/cgn2_1/USP70_spool/US09677752/runat_04052001_132443_4961/app_query.fasta_1.948
-DB=N_Geneseq_0401 -QFMT=fastcap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=4.000 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -START=1 -MATRIX=blosum62
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=25 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09677752 -CGN1_1_255 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-677-752-4
Query length: 878
Database: N_Geneseq_0401.*
Database sequences: 678276
Database length: 291890651
Search time (sec): 148.120000

score_list:

Sequence	Strd Orig	zscore	Escore	Len	Documentation
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A64751		4513.00	6882.52	0.0	2643
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:201425		4395.00	6632.08	0.0	1038602
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A64757		4373.00	6668.22	0.0	2601
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:C81914		947.50	1365.81	3.9e-68	273254
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30852		944.50	1414.69	7.3e-71	2784
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:A30851		944.50	1414.02	8.0e-71	2950
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:A30852		940.50	1408.55	1.6e-70	2787
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:X91990		930.00	1321.46	1.1e-65	1230025
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30853		912.50	1364.79	4.4e-68	3000
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:A30853		906.50	1356.43	1.3e-67	2793
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30854		883.00	1320.34	1.3e-65	2815
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30854		870.00	1300.53	1.7e-64	2790
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30854		866.00	1293.55	4.1e-64	3000
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:C81914		866.00	1240.94	3.5e-61	273254
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:A30859		855.00	1292.19	4.9e-64	2957
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:X06828		857.00	1280.62	2.2e-63	2787
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:X06828		853.50	1274.40	4.8e-63	3000
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:X06828		851.00	1270.38	8.0e-63	3050
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:A30847		849.50	1268.47	1.0e-62	2950
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30847		849.50	1267.52	1.2e-62	3200
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:A30849		847.00	1264.64	1.7e-62	2950
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:X06821		844.50	1261.60	1.5e-62	2957
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:X06827		825.00	1231.38	1.2e-60	2838
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A27342		811.00	1208.72	2.2e-59	3150
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30848		810.00	1208.63	2.2e-59	2784
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30848		786.00	1171.86	2.5e-57	2784
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A48838		773.00	1150.88	3.6e-56	3050
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A47458		763.50	1136.43	2.3e-55	3021
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:X06539		759.50	1125.83	9.0e-55	4435
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:A64752		757.50	1127.52	7.3e-55	2949
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:X06556		753.00	1119.13	2.1e-54	3354
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A50057		749.00	1113.10	4.6e-54	3324
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:A50035		735.00	1093.05	6.1e-53	2950
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:X06818		735.00	1092.65	6.4e-53	3052
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A48839		717.00	1066.04	1.9e-51	2808
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A28691		708.50	1054.26	8.8e-51	2526
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A28690		708.50	1053.26	9.9e-51	2750
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:X06819		705.50	1049.66	1.6e-50	2526
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A28709		574.50	847.49	9.9e-39	2865
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A28708		574.50	846.38	3.3e-39	3150

/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A28715 + 558.50 822.69 6.9e-38 2937
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A28714 + 558.50 821.87 7.7e-38 3150
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A28713 + 536.50 789.55 4.9e-36 2796
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A28712 + 536.50 788.73 5.4e-36 3000
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:261586 + 532.00 781.65 1.3e-35 3050

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A64751

seq_documentation_block:

ID_ A64751 standard; DNA; 2643 bp.
XX AC A64751;
XX DT 02-FEB-2001 (first entry)
XX DE C. trachomatis pmpI gene coding sequence.
XX KW Chlamydial infection; sexually transmitted disease;
KW pelvic inflammatory disease; PID; tubal obstruction; infertility;
KW trachoma; blindness; acute respiratory tract infection;
KW atherosclerosis; coronary heart disease; antibacterial; ss.
XX OS Chlamydia trachomatis.
XX PN WO200034483-A2.
XX PD 15-JUN-2000.
XX PF 08-DEC-1999; 99WO-US29012.
XX PR 08-DEC-1998; 98US-0208277.
PR 08-APR-1999; 99US-0288594.
PR 01-OCT-1999; 99US-0410568.
PR 22-OCT-1999; 99US-0426571.
XX (CORI-) CORIXA CORP.
XX Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;
WPI: 2000-431303/37.
XX Isolated polypeptide for diagnosis and treatment of Chlamydia infection
XX comprises immunogenic portion of Chlamydia antigen, which comprises
XX amino acid sequence encoded by polynucleotide sequence
XX Claim 1: Pages 171-172; 256pp; English.
XX The present invention relates to new nucleic acid sequences and the
XX proteins encoded by the nucleic acid sequences. The encoded proteins
XX comprise an immunogenic portion of a Chlamydia antigen. The encoded
XX proteins are useful for the serodiagnosis and treatment of Chlamydia
XX infection. Chlamydiae are intracellular bacterial pathogens that are
XX responsible for a wide variety of human infections. C. trachomatis
XX infection is one of the most common sexually transmitted diseases and can
XX lead to pelvic inflammatory disease (PID), resulting in tubal obstruction
XX and infertility. Trachoma due to ocular infection with C. trachomatis is
XX the leading cause of preventable blindness worldwide. C. pneumoniae is a
XX major cause of acute respiratory tract infections in humans and is also
XX thought to play a role in the pathogenesis of atherosclerosis and
XX coronary heart disease. The present sequence is a nucleic acid sequence
XX isolated in the present invention.

Sequence 2643 BP; 702 A; 651 C; 503 G; 786 T; 1 other;

alignment_scores:

Quality: 4513.00 Length: 878
Ratio: 5.158 Gaps: 0
Percent Similarity: 99.658 Percent Identity: 99.544

alignment_block:

US-09-677-752-4 x A64751

601	HisGlyGlyAlaTyrLeuPheGlyThrTrpGlySerAlaValSerAsnLeu	617
1807	CACGAGAGAGCCTATCTATTGGAACGTGGGATCTGCTGTTTCTTAATTT	1856
617	uPheTyrAlaHisAspSerGlyLysProIleAspAsnTrpHisHisA	634
1857	ATTCATGTTGTCACGACAGCTCTGGGAAACCTATCGATAATTGGCATCATTA	1906
634	rgSerLeuGlyTyrIleuPheGlyIleSerThrHisSerLeuAspAspHis	650
1907	GAAGCCTTGGCTACCTATTTCGGTATCAGTACTCACAGTTTAGATGACCAT	1956
651	SerPheCysLeuAlaAaGlyGlnLeuLeuGlyLysSerSerAspSerPh	667
1957	TCCTTCTGCTTGGCTGCAGGACAATTACTCGGGAATCGTCCGATTCCTT	2006
667	eIleThrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaGlnL	684
2007	TATTACGCTCACAAACGACCTCCTATATAGCTACTGTACAAGCGCAAC	2056
684	eAlaThrSerLeuMetLysIleSerAlaGlnLacystyrAsnGluSer	700
2057	TCGCTACCTCTCTAATGAAATCTCTGCACAGGCATGCTACAATGAAAGT	2106
701	IleHisGluLeuLysThrLysTyrArgSerPheSerLysGluGlyPheG	717
2107	ATPCCATGAGCTAAAAACAAAATATCGCTCCTTCTTAAGAAGGATTCGG	2156
717	ySerTrpHisSerValAlaValSerGlyGluValCysAlaSerIleProI	734
2157	ATCCYGGCATACCGTTGCAGTATCCGGAGAAGTGTGGCGCATCCATTCTTA	2206
734	IeValSerAsnGlySerGlyLeuPheSerSerPheSerIlePheSerLys	750
2207	TGTATCCAAATGGTTTCGGAGCATGTTTCAGCTCCTTCTCTATTTCTCTAAA	2256
751	LeuGlnGlyPheSerGlyThrGlnAspGlyPheGluSerSerGlyG	767
2257	CTGCAAGGATTTTTCAGGAACACAGACGGTTTTTCAGGAGAGTTCGGGAGA	2306
767	uIleArgSerPheSerAlaSerSerPheArgAsnIleSerLeuProIleG	784
2307	GATTTCGGTCTTTCTGCCAGCTCTTTTCAGAAATATTCTACTTCCTATAG	2356
784	LyIleThrPheGluLysIysSerGlnLysThrArgThrTyrTyrTyrPhe	800
2357	GAATATACATTTGAAAAAATCCCAAAACACAGACCTACTATTACTTT	2406
801	LeuGlyAlaTyrIleGlnAspLeuLysArgAspValGluSerGlyProVa	817
2407	CTAGAGCCTACATCCAGACCTGAAACGTGATGTGGATCGGGACCTGT	2456
817	IValLeuLeuLysAsnAlaValSerTyrPaspAlaProMetAlaAsnLeuA	834
2457	AGTGTACTCAAAATGCGCTCTCTGGGATGCTCCTATGCGCAACTTGG	2506
834	spSerArgAlaTyrMetPheArgLeuThrAsnGlnArgAlaLeuHisArg	850
2507	ATTCACGAGCCTACATGTTCCGGCTTACGAATCAAGAGCTCTACACAGA	2556
851	LeuGlnThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSe	867
2557	CTTCAGACGCTGTTAAATGTGCTTGTGTGCTGCGTGGGCNAACCATAG	2606
867	rTyrSerLeuAspLeuGlyThrThrTyrArgPhe	878
2607	TTACTCCCTGGATCTGGGACCACTTACAGGTTTC	2640

seq_name: /SIDS2/qcdata/geneseq/geneseq/NA1999.DAT:Z01425

seq_documentation_block:

```
seq_documentation_block:
ID      201425 standard; DNA: 1038602 BP.
```

XX
XX

AC	Z01425;
XX	
XX	07-OCT-1999 (first entry)
XX	
DE	Complete genome sequence of Chlamydia trachomatis.
XX	
KW	Vaccine: eye disease; conventional trachoma; nonendemic trachoma;
KW	paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;
KW	nongonococcal urethritis; epidymitis; cervicitis; salpingitis;
KW	bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
XX	
OS	Chlamydia trachomatis.
XX	
XX	W09928475-A2.
PN	
XX	10-JUN-1999.
PD	
XX	27-NOV-1998; 98WO-1B01939.
XX	
XX	04-NOV-1998; 98US-0107077.
PR	28-NOV-1997; 97FR-0015041.
PR	17-DEC-1997; 97FR-0016034.
XX	
XX	(G8ST) GENSET.
PA	
XX	
PI	Griffais R;
XX	
XX	WPI: 1999-371125/31.
DR	
XX	Genome sequence of Chlamydia trachomatis
PT	
XX	Claim 1; Page 373-656; 1755pp; English.
PS	
XX	
CC	The present sequence represents the complete genome of Chlamydia
CC	trachomatis. Open reading frames (ORFs) of the genome encode polypeptides
CC	Y36754-Y37949. The polypeptides can be used as vaccines against
CC	Chlamydia trachomatis. Antisense and ribozyme sequences can also be
CC	used to control growth of the microorganism. Chlamydia trachomatis is
CC	responsible for a large number of diseases, e.g. eye diseases such as
CC	conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion
CC	conjunctivitis; genital diseases such as nongonococcal urethritis,
CC	epidymitis, cervicitis, salpingitis, perihhepatitis, bartholinitis;
CC	pneumopathy in breast feeding infants; and venereal lymphogranulomatosis.
CC	The polypeptides of the invention may be of use in treating these
CC	diseases.
XX	
SO	Sequence 1038602 BP: 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

```

667 heileThrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaGln 683
|||||
473862 TTATTAGCTCAGAAACACGCTCTTATATAGTACTGTACAGCGCAA 473911
|||||
684 LeuAlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrAsnGluSe 700
|||||
473912 CTCGCTACCTCTCTATGAATACTCTGCACAGCGCATGCTACAATGAAG 473961
|||||
700 rIleHisGluLeuLysThrLysTyrArg.SerPheSerLysGluGlyPhe 716
|||||
473962 TATCCATGAGCTAAACAAATATCGCTCTCTCTAAAGAGGATTC 474011
|||||
717 GlySerTrpHisSerValAlaValSerGlyGluValCysAlaSerIlePr 733
|||||
474012 GGATCCCTGGCATGCGTTGAGATATCCGAGAGAGTGTGCGCATCGATCC 474061
|||||
733 oIleValSerAsnGlySerGlyLeuPheSerSerPheSerIlePheSerL 750
|||||
474062 TATTGTATCAATGCTTCGCGACTGTTACGCTCTCTCTATTTCTCTA 474111
|||||
750 ysLeuGlnGlyPheSerGlyThrGlnAspGlyPheGluGluSerSerGly 766
|||||
474112 AACTGCAAGGATTTTCAGAACACAGACGCGTTTGGAGAGAGTTCGGA 474161
|||||
767 GluIleArgSerPheSerAlaSerSerPheArgAsnIleSerLeuProII 783
|||||
474162 GAGATTCGGTCCCTTCTGCGAGCTCTTTTCAGAAATATTTCACTTCCAT 474211
|||||
783 eGlyIleThrPheGluLysLysSerGlnLysThrArgThrTyrTyrTrp 800
|||||
474212 AGGAATAACATTTGAAAAAATCCCAAAACACAGCACTACTATTACT 474261
|||||
800 heLeuGlyAlaTyrIleGlnAspLeuLysArgAspValGluSerGlyPro 816
|||||
474262 TTCTAGAGCGCTACATCAAGACCTGAAGCTGATGTGGAATCGGGACCT 474311
|||||
817 ValValLeuLeuLysAsnAlaValSerTrpAspAlaProMetAlaAsnLe 833
|||||
474312 GTAGTGTTACTCAAAATGCGCTCTCTGGGATGCTCCTATGGCGAACTT 474361
|||||
833 uAspSerArgAlaTyrMetPheArgLeuThrAsnGlnArgAlaLeuHisA 850
|||||
474362 GGATTCAGAGCGCTACATGTTCCGGCTTACGAATCAAGAGCTCTACACA 474411
|||||
850 rgLeuGlnThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHis 866
|||||
474412 GACTTCAGAGCGCTTAAATGTGTCTTGTGTGCTGCTGGCGCAAG.CAT 474460
|||||
867 SerTyrSerLeuAspLeuGlyThr 874
|||||
474461 AGTTACTCCCTGGATCTGGGGACA 474484
|||||

```

seq_name: /SDS2/gcgdata/geneseq/geneseq/NA2000.DAT.A64757

seq_documentation_block:

ID A64757 standard; DNA: 2601 BP.

XX AC A64757;

XX DT 02-FEB-2001 (first entry)

DE C. trachomatis pmpI gene coding sequence minus the signal sequence.

XX Chlamydial infection; sexually transmitted disease;

KW pelvic inflammatory disease; PID; tubal obstruction; infertility;

KW trachoma; blindness; acute respiratory tract infection;

KW atherosclerosis; coronary heart disease; antibacterial; ss.

XX Chlamydia trachomatis.

OS Chlamydia trachomatis.

XX WO200034483-A2.

XX PN 15-JUN-2000.

XX PD

```

XX 08-DEC-1999: 99W0-US29012.
XX 08-DEC-1998: 98U5-0208277.
XX 08-APR-1999: 99U5-0288594.
XX 01-OCT-1999: 99U5-0410568.
XX 22-OCT-1999: 99U5-0426571.
XX (CORI-) CORIXA CORP.
XX Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;
XX WPI: 2000-431303/37.

```

Isolated polypeptide for diagnosis and treatment of Chlamydia infection comprises immunogenic portion of Chlamydia antigen, which comprises amino acid sequence encoded by polynucleotide sequence

Claim 1: Pages 198-199; 256pp; English.

The present invention relates to new nucleic acid sequences and the proteins encoded by the nucleic acid sequences. The encoded proteins comprise an immunogenic portion of a Chlamydia antigen. The encoded proteins are useful for the serodiagnosis and treatment of Chlamydia infection. Chlamydiae are intracellular bacterial pathogens that are responsible for a wide variety of human infections. C. trachomatis infection is one of the most common sexually transmitted diseases and can lead to pelvic inflammatory disease (PID), resulting in tubal obstruction and infertility. Trachoma due to ocular infection with C. trachomatis is the leading cause of preventable blindness worldwide. C. pneumoniae is a major cause of acute respiratory tract infections in humans and is also thought to play a role in the pathogenesis of atherosclerosis and coronary heart disease. The present sequence is a nucleic acid sequence isolated in the present invention.

Sequence 2601 BP; 696 A; 641 C; 490 G; 769 T; 5 other;

alignment_scores:

Quality: 4373.00 Length: 857
Ratio: 5.127 Gaps: 0
Percent Similarity: 99.533 Percent Identity: 98.950

alignment_block:

US-09-677-752-4 x A64757

Align seg 1/1 to: A64757 from: 1 to: 2601

```

22 leupheGlyGlnAspProLeuGlyGluThrAlaLeuLeuThrLysAsnPr 38
|||||
28 CTCTTTGGCCAGGATCCCTTAGGTGAACCCGCCCTCCTCACTAAAAATCC 77
|||||
38 oAsnHisValValCysThrPhePheGluAspCysThrMetGluSerLeup 55
|||||
78 TAATCATGCTGCTGTACATTTTTTGGAGGACTGTACCATGGAGAGCCTCT 127
|||||
55 heProAlaLeuCysAlaHisAlaSerGlnAspAspProLeuTyrValLeu 71
|||||
128 TTCCTGCTCTTTGTGCTCATGCATCAAGACGATCCTTTGTATGTACTT 177
|||||
72 GlyAsnSerTyrCysTrpPheValSerLysLeuHisIleThrAspProLy 88
|||||
178 GGAAATTCCTACTCTTGGTTCGTATCTAAACTCCATATCAGGACCCCAA 227
|||||
88 sGluAlaLeuPheLysGluLysGlyAspLeuSerIleGlnAsnPheArgp 105
|||||
228 AGAGCTCTTTTAAAGAAAAAGAGATCTTTCCATTCAAAACCTTCGCT 277
|||||
105 heLeuSerPheThrAspCysSerSerLysGluSerGluSerProSerIlele 121
|||||
278 TCCCTTCTTCACAGATTGCTCTTCCCAAGGAAGAGCTCTCCTCTTATTATT 327
|||||
122 HisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPh 138
|||||

```

738 ySerGlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGlyPheS 755
 |||||
 2178 TTCGGAGCTGTTACGTCCTCTCTATTTCTCTAAATCTCAAGGATTTT 2227
 |||||
 755 erGlyThrGlnAspGlyPheGluGluSerSerGlyGluIleArgSerPhe 771
 |||||
 2228 CAGGAACACAGGAGCGTTTTCAGGAGAGTTTCGGGAGAGATTCGGTCCTTT 2277
 |||||
 772 SerAlaSerSerPheArgAsnIleSerLeuProIleGlyIleThrPheG1 788
 |||||
 2278 TCTCCAGCTCTTTTCAGAAATATTTCTACTTCTCTATAGGAATAACATTGA 2327
 |||||
 788 uLysLysSerGlnLysThrArgThrTyrTyrTyrPheLeuGlyAlaTyrI 805
 |||||
 2328 AAAAAATCCCAAAAAACACGAACTACTATTACTTTCTAGGAGCGCTACA 2377
 |||||
 805 leGlnAspLeuLysArgAspValGluSerGlyProValValLeuLys 821
 |||||
 2378 TCCAAAGACCTGAACAGTGATGTGAATCGGAGACCTGTAGTGTACTCAA 2427
 |||||
 822 AsnAlaValSerTrpAspAlaProMetAlaAsnLeuAspSerArgAlaTy 838
 |||||
 2428 AATCCGCTCTCTGGGATGCTCTCTATGCGGAACCTGGATTCCAGGCGCTA 2477
 |||||
 838 rMetPheArgLeuThrAsnGlnArgAlaLeuHisArgLeuGlnThrLeuL 855
 |||||
 2478 CATGTCGGGCTTACGAATCAAGAGCTCTACACAGACTTCAGACGCTGT 2527
 |||||
 855 euAsnValSerCysValLeuArgGlyGlnSerHisSerTyrSerLeuAsp 871
 |||||
 2528 TAAATGTCCTGTGTGCTGCTGGCGGAACCCATAGTTACTCTCCTGGAT 2577
 |||||
 872 LeuGlyThrThrTyrArgPhe 878
 |||||
 2578 CTGGGACCACTTACAGGTTT 2598

seq_name: /SDS2/gcgdata/geneseq/geneseq/NA2000.DAT:C81914

seq_documentation_block:

ID C81914 standard; DNA; 273254 BP.

AC C81914;

XX 27-FEB-2001 (first entry)

XX Chlamydia pneumoniae genome DNA.

XX Genome; diagnosis; vaccine; ds.

XX Chlamydia pneumoniae.

XX WO200027994-A2.

XX 18-MAY-2000.

XX 12-NOV-1999; 99WO-US26923.

XX 12-NOV-1998; 98US-0108279.

XX 08-APR-1999; 99US-0128606.

XX (REGC) UNIV CALIFORNIA.

XX Stephens R, Mitchell W, Kalman S, Davis R;

XX WPI: 2000-376516/32.

XX Isolated nucleic acid for use in diagnostic and analytical methods
 PT encodes genomic sequence of Chlamydia pneumoniae -

XX Claim 2: Page 128-320; 320pp; English.

XX This invention describes a novel nucleic acid (N1) encoding a Chlamydia

CC pneumoniae protein (P1), given in the specification. The isolated nucleic
 CC acid is useful for diagnostic and analytical methods. The protein may
 CC hybridization-based assays or amplification-based assays. The protein may
 CC be used for diagnostic purposes, for their enzymatic or structural
 CC activity, or as a vaccine. The invention also describes (1) a probe
 CC comprising a hybridizing fragment of N1; (2) an isolated nucleic acid
 CC (N2) that hybridizes under stringent conditions to N1; (3) an expression
 CC cassette comprising N1 under the transcriptional regulation of a
 CC transcriptional initiation region functional in an expression host, and a
 CC transcriptional termination region; (4) a cell comprising an expression
 CC cassette of (3) as part of an extrachromosomal element or integrated into
 CC the genome of a host cell as a result of induction of the expression
 CC cassette into the host cell, and the cellular progeny of the host cell;
 CC (5) a method for producing a P1 comprising growing a cell of (4) where
 CC the protein is expressed and isolating the protein free of other
 CC proteins; (6) a purified polypeptide composition comprising at least 50
 CC weight % of P1; and (7) a monoclonal antibody binding specifically to the
 CC peptide of (6).
 XX
 SQ Sequence 273254 BP; 76423 A; 51054 C; 61965 G; 83812 T; 0 other;

alignment_scores:

Quality: 947.50 Length: 910

Ratio: 1.771 Gaps: 32

Percent Similarity: 58.791 Percent Identity: 31.648

alignment_block:

US-09-677-752-4 x C81914/rev ..

Align seg 1/1 to reverse of: C81914 from: 1 to: 273254

85 ThrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerIleG1 101
 |||||
 105585 ACAATGCAGATGGAAGTATCTATATCTTAACAGGGGATGCTCAATCAC 105536
 |||||
 101 nAsnPheArgPhe...LeuSerPheThrAspCysSerSerLysGluSerS 117
 |||||
 105535 CAATCGAGGATCTCCGACAGCTCTAACCGCTTCTGCTGCTTTAAAGAACTA 105486
 |||||
 117 erProSerIleIleHisGlnLysAsnGly...GlnLeuSerLeuArgAsn 132
 |||||
 105485 CTGGGAATCTTTCTTCCAGGCCACGCTACCAATTTCTCTACAAAT 105436
 |||||
 133AsnGlySe 135
 |||||
 105435 ATCGATCGGGAGCGAACTGTACCTTTACCAATACAGCTGCAATAAGCT 105386
 |||||
 135 rMetSerPhe.....CysArgAsnH 142
 |||||
 105385 TCTCTCTTTTCAGGATTCCTCTATTGTCATAATACAAACCACGCAATG 105336
 |||||
 142 isAlaGluGlySerGlyAlaIleSerAlaAspAlaPheSerLeuGln 158
 |||||
 105335 CTACCACAGGAACAGAGCCCATCAAGTCCACAGGAGCTGTGCTTCTTCAG 105286
 |||||
 159 HisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerLysGlyAsnG1 175
 |||||
 105285 TCGAACTATV...ACTTGCTACTTTTGGCCAAACTTTTCTAATGACAATGG 105239
 |||||
 175 yGlyAlaIleGlnAlaGlnThrPheSerLeuSerArgAsnValSerProI 192
 |||||
 105238 AGCGCCCTCCAAAGGCGACTCTATCAGTCTATCGCTAAACCCCAAC...C 105192
 |||||
 192 leSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCysCys 208
 |||||
 105191 TAAGCTTTGCCAAAACAAACACGCAACGCAAAAGGGGTGCTCTATTCC 105142
 |||||
 209 SerAsnLeuIleCysSerGlyAsn...ValAsnProLeuPhePheThrG1 224
 |||||
 105141 ACGGAGGGATTACAATTAAACAATACGTTAAACTCAGCATCATTTCTCTGA 105092
 |||||
 224 yAsnSerAlaThrAsn..... 229

```

761 heGluGlu...SerSerGlyGluIleArgSerPheSerAlaSerSerPhe 776
   ||:|||||  ||:|||||  ||:|||||  ||:|||||  ||:|||||
103313 TCAACAACGTAATACTACTCTGTCAGCATCTTTCGATACGGGTGATTGA 103264

777 ArgAsnIleSerLeuProIleGlyIleThrPheGluLysLysSerGlnly 793
   ||:|||||  ||:|||||  ||:|||||  ||:|||||  ||:|||||
103263 ATTACGCTCTCTGCTGCTATTGGAATACCTTCGAGAGATTCTCGAGAA 103214

793 sThrArgThrTyrTyrTyPheLeuGlyAlaTyrIleGlnAspLeuLysA 810
   ||:|||||  ||:|||||  ||:|||||  ||:|||||  ||:|||||
103213 CGACGCTGGCTTACGAGCTACTCTCATCTACGTTCGGGTGCTATC 103164

810 rgAspValGluSerGlyProValValLeuLeuLysAsnAlaValSerTrp 826
   ||:|||||  ||:|||||  ||:|||||  ||:|||||  ||:|||||
103163 GTAAGAATCTGACTCCAGCACAGCTCTCTAATCAACAATACCTCGTGG 103114

827 AspAlaProMetAlaAsnLeuAspSerArgAla..... 837
   ||:|||||  ||:|||||  ||:|||||  ||:|||||  ||:|||||
103113 AAAACTACAGGAACGAACTCTCTCAAGACAAAGCTGTGTCGGAAGACAGG 103064

838 .....TyrMetPheArgLeuThrAsnGlnArgAlaLeuHisArgLeuG 852
   ||:|||||  ||:|||||  ||:|||||  ||:|||||  ||:|||||
103063 GATCTTTATGCTCTCT.....CCNAATCTTG 103035

852 lnThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSerTyr 868
   ||:|||||  ||:|||||  ||:|||||  ||:|||||  ||:|||||
103034 AGCTCAAGTAACCTATCTATGAAATTCGTGGATCTTCACGACGCTAC 102985

869 SerLeuAspLeuGlyThrThrTyrArgPhe 878
   ||:|||||  ||:|||||  ||:|||||  ||:|||||  ||:|||||
102984 AATGCAGATCTTGGAGGTAAATTCAGTTC 102955

```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A30852

seq_documentation_block:

ID A30852 standard; DNA: 2784 BP.

XX AC A30852;

XX DT 29-AUG-2000 (first entry)

XX DE Chlamydia antigen CPN100638 gene open reading frame.

XX KW Chlamydia antigen; diagnosis: infection; community acquired pneumonia;
 KW therapy; upper respiratory tract disease; bronchitis; sinusitis;
 KW asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma;
 KW ds.

XX OS Chlamydia pneumoniae.

XX FH Key Location/Qualifiers

XX CDS 1..2784

XX FT /*tag= a

XX FT /partial

XX FT /product= chlamydia antigen CPN100638

XX FT /note= "no stop codon given"

XX PN WO200032794-A2.

XX PD 08-JUN-2000.

XX PF 01-DEC-1999; 99WO-CA01147.

XX PR 01-DEC-1998; 98US-0110339.

XX PR 01-DEC-1998; 98US-0110340.

XX PR 01-DEC-1998; 98US-0110427.

XX PR 01-DEC-1998; 98US-0110428.

XX PR 01-DEC-1998; 98US-0110438.

XX PA (CONN-) CONNAUGHT LAB LTD.

XX XX Murdin AD, Oomen RP, Wang J;

PI

XX DR

WPI: 2000-412339/35.

P-PSDB: Y90239.

XX XX

PT Nucleic acids encoding polypeptide antigens from Chlamydia useful for preventing, diagnosing and treating diseases such as community acquired pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset asthma.

XX PT

Claim 2; Fig 5; 174pp; English.

XX PS

CC This sequence encodes a Chlamydia antigen of the invention, designated CPN100638. The nucleic acids (and their complementary sequences) may be used as diagnostic agents for detecting the presence of nucleic acids encoding Chlamydia antigens in samples according to standard methods, and therefore, for diagnosing Chlamydia infections. For example, they may be used as primers and probes for diagnostic polymerase chain reaction (PCR) assays. Antisense sequences may be used to down regulate expression of the proteins and may be used to treat infections. The nucleic acids may also be used to produce the protein antigens they encode according to standard recombinant DNA methodologies. The proteins may then be used as antigens for the production of antibodies (i.e. as vaccines) for preventing infection by Chlamydia. The antibodies may also be used as diagnostic reagents for detecting infections. Chlamydia is a pathogen implicated in the development of (for example) community acquired pneumonia, upper respiratory tract disease (especially bronchitis and sinusitis), asthmatic bronchitis, adult-onset asthma and acute exacerbations of asthma in adults.

XX SQ

Sequence 2784 BP; 811 A; 689 C; 537 G; 747 T; 0 other:

alignment_scores:

Quality: 944.50 Length: 910

Ratio: 1.769 Gaps: 32

Percent Similarity: 58.681 Percent Identity: 31.648

alignment_block:

US-09-677-752-4 x A30852

Align seg 1/1 to: A30852 from: 1 to: 2784

85 ThrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerIleG1 101

154 ACAATGCCAGATGCAACTATCTATAATCTAACAGGGGATGCTCTCAATCAC 203

101 nAsnPheArgPhe....LeuSerPheThrAspCysSerSerLysGluSers 117

204 CAATGCAGGATCTCCGACAGCTCTAACCGCTTCCTGCTTTTAAAGAACTA 253

117 erProSerIleIleHisGlnLysAsnGly....GlnLeuSerLeuArgAsn 132

254 CTGGGAATCTTTCTTCCAAAGGCCACGGCTACCAATTTCTCTACAAAT 303

133AsnGlyse 135

304 ATCGATCGGGGCGGAAGCTTACCTTTTACCAATACAGCTGCAATAAGCT 353

135 rMetSerPhe.....CysArgAsnH 142

354 TCTCTCTTTTCAGGATTCCTCTATTGTCACCTAATACAAACCCAGCAATG 403

142 isAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSerLeuGln 158

404 CTACCACAGGAACAGAGCCATCAAGTCCACAGGAGCTTGTCTATTTCAG 453

159 HisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerLysGlyAsnG1 175

454 TCGAACTAT...AGTTGCTACTTTGGCCAAAACCTTTTCTAATGACAATGG 500

175 yGlyAlaIleGlnAlaGlnThrPheSerLeuSerArgAsnValSerPro1 192

501 AGCGGCCCTCCAAAGGAGCTCTTATCATGCTCTATCGCTAAACCCCAAC...C 547

PR 01-DEC-1998; 98US-0110340.
PR 01-DEC-1998; 98US-0110427.
PR 01-DEC-1998; 98US-0110428.
PR 01-DEC-1998; 98US-0110438.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Murdin AD, Oomen RP, Wang J;
XX WPI: 2000-412339/35.
XX P-PSDB; Y90239.
XX
XX Nucleic acids encoding polypeptide antigens from Chlamydia useful for
PT preventing, diagnosing and treating diseases such as community acquired
PT pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset
PT asthma -
XX
XX Claim 2; Fig 5; 174pp; English.
XX
XX This sequence encodes a Chlamydia antigen of the invention, designated
CC CPN100638. The nucleic acids (and their complementary sequences) may be
CC used as diagnostic agents for detecting the presence of nucleic acids
CC encoding Chlamydia antigens in samples according to standard methods,
CC and therefore, for diagnosing Chlamydia infections. For example, they may
CC be used as primers and probes for diagnostic polymerase chain reaction
CC (PCR) assays. Antisense sequences may be used to down regulate
CC expression of the proteins and may be used to treat infections. The
CC nucleic acids may also be used to produce the protein antigens they
CC encode according to standard recombinant DNA methodologies. The
CC proteins may then be used as antigens for the production of antibodies
CC (i.e. as vaccines) for preventing infection by Chlamydia. The
CC antibodies may also be used as diagnostic reagents for detecting
CC infections. Chlamydia is a pathogen implicated in the development of
CC disease (especially bronchitis and sinusitis, upper respiratory tract
CC adult-onset asthma and acute exacerbations of asthma in adults.
XX
XX Sequence 2950 BP; 855 A; 724 C; 562 G; 809 T; 0 other;

alignment_scores:
Quality: 944.50 Length: 910
Ratio: 1.769 Gaps: 32
Percent Similarity: 58.681 Percent Identity: 31.648

alignment_block:
US-09-677-752-4 x A30851

Align seg 1/1 to: A30851 from: 1 to: 2950

85 ThrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerIleG 101
|||
254 ACAAAATGCAGATGGAACATATCTATAATCTAACAGGGGATGCTCAATCAC 303
|||
101 nAsnPheArgPhe...LeuSerPheThrAspCysSerSerLysGluSerS 117
|||
304 CAATGCAGATCTCCGACAGCTCTACCCCTCTCTCTTTAAAGAACTA 353
|||
117 erProSerIleLeHisGlnLysAsnGly...GlnLeuSerLeuArgAsn 132
|||
354 CTGGGAATCTTTCTTCAAGGCCAGCGGTACCAATTTCTCTACAAAT 403
|||
133AsnGlySer 135
|||
404 ATCGATGCGGGAGCGAACTGTACCTTTACCAATACAGCTGCCAATAGCT 453
|||
135 rMeLSerPhe.....CysArgAsnH 142
|||
454 TCCTCTCTTTTTCAGGATTCCTCTATTGTCCTAATAACAAACCACGAATG 503
|||
142 isAlaGluGlySerGlyAlaIleSerAlaAspAlaPheSerLeuGln 158
|||
504 CTACCACAGGACAGGAGCCATCAAGTCCACAGAGCTTGTCTATTTCAG 553
|||

PR 01-DEC-1998; 98US-0110340.
PR 01-DEC-1998; 98US-0110427.
PR 01-DEC-1998; 98US-0110428.
PR 01-DEC-1998; 98US-0110438.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Murdin AD, Oomen RP, Wang J;
XX WPI: 2000-412339/35.
XX P-PSDB; Y90239.
XX
XX Nucleic acids encoding polypeptide antigens from Chlamydia useful for
PT preventing, diagnosing and treating diseases such as community acquired
PT pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset
PT asthma -
XX
XX Claim 2; Fig 5; 174pp; English.
XX
XX This sequence encodes a Chlamydia antigen of the invention, designated
CC CPN100638. The nucleic acids (and their complementary sequences) may be
CC used as diagnostic agents for detecting the presence of nucleic acids
CC encoding Chlamydia antigens in samples according to standard methods,
CC and therefore, for diagnosing Chlamydia infections. For example, they may
CC be used as primers and probes for diagnostic polymerase chain reaction
CC (PCR) assays. Antisense sequences may be used to down regulate
CC expression of the proteins and may be used to treat infections. The
CC nucleic acids may also be used to produce the protein antigens they
CC encode according to standard recombinant DNA methodologies. The
CC proteins may then be used as antigens for the production of antibodies
CC (i.e. as vaccines) for preventing infection by Chlamydia. The
CC antibodies may also be used as diagnostic reagents for detecting
CC infections. Chlamydia is a pathogen implicated in the development of
CC disease (especially bronchitis and sinusitis, upper respiratory tract
CC adult-onset asthma and acute exacerbations of asthma in adults.
XX
XX Sequence 2950 BP; 855 A; 724 C; 562 G; 809 T; 0 other;

alignment_scores:
Quality: 944.50 Length: 910
Ratio: 1.769 Gaps: 32
Percent Similarity: 58.681 Percent Identity: 31.648

alignment_block:
US-09-677-752-4 x A30851

Align seg 1/1 to: A30851 from: 1 to: 2950

85 ThrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerIleG 101
|||
254 ACAAAATGCAGATGGAACATATCTATAATCTAACAGGGGATGCTCAATCAC 303
|||
101 nAsnPheArgPhe...LeuSerPheThrAspCysSerSerLysGluSerS 117
|||
304 CAATGCAGATCTCCGACAGCTCTACCCCTCTCTCTTTAAAGAACTA 353
|||
117 erProSerIleLeHisGlnLysAsnGly...GlnLeuSerLeuArgAsn 132
|||
354 CTGGGAATCTTTCTTCAAGGCCAGCGGTACCAATTTCTCTACAAAT 403
|||
133AsnGlySer 135
|||
404 ATCGATGCGGGAGCGAACTGTACCTTTACCAATACAGCTGCCAATAGCT 453
|||
135 rMeLSerPhe.....CysArgAsnH 142
|||
454 TCCTCTCTTTTTCAGGATTCCTCTATTGTCCTAATAACAAACCACGAATG 503
|||
142 isAlaGluGlySerGlyAlaIleSerAlaAspAlaPheSerLeuGln 158
|||
504 CTACCACAGGACAGGAGCCATCAAGTCCACAGAGCTTGTCTATTTCAG 553
|||

```
2279 TACCTTCTGGATCTGAAGTGACGCGCTCTCTCTTTGATGCTCAGAT 2328
      ::::::::::::::
695 acyTyrAsnGluSerIleHisGluLeuLysThrLysTyrArgSerPheS 712
      ||| :::::::::::::: |||
2329 CAGCTATATCTATGTAATACTATGAAACCTATTACACCCCAAGCAC 2378
      ||| :::::::::::::: |||
712 erLysGluGlyPheGlySerTrpHisSerValAlaValSerGlyGluVal 728
      ||| :::::::::::::: |||
2379 CAAG...GGAGAGAGCTCGTGGTATATGACGCTCGCTCTGGAACCT 2425
      ::::::::::::::
729 CysAlaSerIleProIleValSer...AsnGlySerGlyLeuPheSerSe 744
      ::::::::::::::
2426 CGGAGCTCCCTACACACACACGCTTTTAAAGCATGAGGCTCTCTCCACGC 2475
      ::::::::::::::
744 rPheSerIlePheSerLysLeuGlnGlyPheSerGlyThrGlnAspGlyP 761
      ::::::::::::::
2476 GTATTTCTCTTTCATCAAGTAGAAGCTTCGTACATACACCAAGATAGCT 2525
      ::::::::::::::
761 heGluGlu...SerSerGlyGluLeuArgSerPheSerAlaSerSerPhe 776
      ||| :::::::::::::: |||
2526 TCAAGAACGTAATCTACTACCTTGTGATGATCTTCGATAGCGGTGATTA 2575
      ||| :::::::::::::: |||
777 ArgAsnIleSerLeuProIleGlyIleThrPheGluLysLysSerGlnLy 793
      ||| :::::::::::::: |||
2576 ATTACGCTCTGTGCTTATGGAAATACCTTCGAGAGATCTCGAGNAA 2625
      ||| :::::::::::::: |||
793 sThrArgThrTyrTyrPheLeuGlyAlaTyrIleGlnAspLeuLysA 810
      ||| :::::::::::::: |||
2626 CGAGCGTGGCTTTACGAAGCTACTGTCTATCTAGCTTCCGATGCTATC 2675
      ||| :::::::::::::: |||
810 rGaspValGluSerGlyProValValLeuLeuLysAsnAlaValSerTrp 826
      ||| :::::::::::::: |||
2676 GTAAGAATCTCTGACTGCGACGACGCTCTCTCAATCAACAATACCTCGTG 2725
      ::::::::::::::
827 AspAlaProMetAlaAsnLeuAspSerArgAla..... 837
      ::::::::::::::
2726 AAACTACAGAGAGATCTCTCAAGACAGCTGTATCGGAAGACGAG 2775
      ::::::::::::::
838 .....TyrMetPheArgLeuThrAsnGlnArgAlaLeuHisArgLeuG 852
      ||| |||
2776 GATCTTTTATGCTTCTCT.....CCAAATCTTG 2804
      ::::::::::::::
852 lnThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSerTyr 868
      ::::::::::::::
2805 AGGTCACAAAGTAACCTATCTATGAAATTCGTGGATCTTCACGACGTAC 2854
      ::::::::::::::
869 SerLeuAspLeuGlyThrThrTyrArgPhe 878
      ::::::::::::::
2855 AATGCAGATCTTGGAGGTAAAGTTCAGTTC 2884
      ::::::::::::::
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:X06822

seq_documentation_block:

ID X06822 standard; DNA; 2787 BP.

XX

AC X06822;

XX

DT 26-APR-1999 (first entry)

XX

DE Chlamydia pneumoniae surface exposed protein Omp10 DNA.

XX

KW Omp10; outer membrane protein 10; surface exposed protein; antigen;

XX

OS Chlamydia pneumoniae.

XX

PN W09858953-A2.

XX

PD 30-DEC-1998.

XX

PF 19-JUN-1998; 98WO-DK00266.

XX

```
PR 23-JUN-1997; 97DK-0000744.
XX (BIRK/) BIRKELUND S.
PA (CHRI/) CHRISTIANSEN G.
XX
PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI Mygind P;
XX
DR WPI: 1999-105610/09.
DR P-PSDB: W88423.
XX
PT Species-specific test for identifying mammals infected with
PT Chlamydia pneumoniae - comprises detecting antibodies specific for
PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
PT these proteins
XX
PS Claim 6; Page 59; 115pp; English.
XX
CC This DNA sequence codes for the novel 98.4 kDa surface exposed
CC protein Omp10 (see W88423) of the human respiratory pathogen
CC Chlamydia pneumoniae. By generating antibodies against C.
CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
CC was obtained which reacted with outer membrane proteins. The
CC antibody was used to identify the genes (see X06816-27) encoding
CC Omp4-Omp15 proteins (see W88417-28) in an expression library of
CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:
CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
CC the other, and encode polypeptides of about 89.6-100.3 kDa and
CC about 56.1 kDa. The invention provides a new species specific test
CC for identifying mammals (including humans) infected with Chlamydia
CC pneumoniae. The test comprises detecting antibodies specific for
CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used
CC in the diagnosis of C. pneumoniae infection in mammals. The
CC nucleic acids and proteins can also be used in the immunization of
CC mammals, the nucleic acids being particularly useful as DNA
CC vaccines for effecting in vivo expression of antigens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with C. pneumoniae.
XX
SQ Sequence 2787 BP; 815 A; 689 C; 535 G; 748 T; 0 other;
```

alignment_scores:

Quality: 940.50 Length: 910

Ratio: 1.761 Gaps: 32

Percent Similarity: 58.681 Percent Identity: 31.538

alignment_block:

US-09-677-752-4 x X06822 ..

Align seg 1/1 to: X06822 from: 1 to: 2787

```
85 ThrAspProLysGluAlaLeuPheLysGlyLysGlyAspLeuSerIleG1 101
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
154 ACAATGCGAGATGGAACATCTATACTAACAGGGGATGCTCAATCAC 203
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
101 nAsnPheArgPhe...LeuSerPheThrAspCysSerSerLysGluSers 117
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
204 CAATGCGAGGATCTCCGACAGCTTAACCGCTTCTCTGCTTTAAAGAACTA 253
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
117 erProSerIleIleHisGlnLysAsnGly...GlnLeuSerLeuArgAsn 132
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
254 CTGGAATCTTTCTTTCCAGGCCACGCTACCAATTTCTCTACAAAT 303
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
133 .....AsnGlySe 135
|||
304 ATCGATGCGGGAGCGAAGTGTACCTTTTACCANTACAGCTGCNAATAAGCT 353
|||
135 rMetSerPhe.....CysArgAsnH 142
:::|:::|:::|
354 TCTCTCTTTTCAGGATCTCTCTATTTTGTCACTAATACAAACACGAAATG 403
|||
```


2129 CTCTCCATCTCCAGCATCTAGCAGCTTGTCTCTCCAAAGCTTGTACGC 2178
 691IleSerAlaGlnAl 695
 2179 TACCTTCTCGATCTGAAGTGAAGCAGCCTCTCTCTTTGTATGTCAGAT 2228
 695 aCysTyrAsnGluSerIleHisGluLeuLysThrLysTyrArgSerPheS 712
 2229 CAGCTATATCTATAGTAATAATACTATGAAACCTATTACACCCCAAGC 2278
 712 eRLysGluGlyPheGlySerTyrHisSerValAlaValSerGlyGluVal 728
 2279 CAAG...GGAGAGCTCGTGGTATATGACGGTTGCGCTCTGGAACVT 2325
 729 CysAlaSerIleProIleValSer...AsnGlySerGlyLeuPheSerSe 744
 2326 GCGAGCTCCCTACCACACACTGCTTTAAGCCATGAGGCTCTCTCCAGC 2375
 744 rPheSerIlePheSerLysLeuGlnGlyPheSerGlyThrGlnAspGlyP 761
 2376 GTATTTCTCTTCATCAAGTAGAAGCTTCGTACATACACACAGATAGCT 2425
 761 heGluGlu...SerSerGlyGluIleArgSerPheSerAlaSerSerPhe 776
 2426 TCAAGAACGTAATACTACTCTGGTACCATCTTTCGATAGCGGTGATTTA 2475
 777 ArgAsnIleSerLeuProIleGlyIleThrPheGluLysLysSerGlnLys 793
 2476 ATTAACGCTCTGTGCTTATTGGAATTACCTTCGAGAGATTCTCGAGAA 2525
 793 sThrArgThrTyrTyrTyrPheLeuGlyAlaTyrIleGlnAspLeuLysA 810
 2526 CGAGCGTGGCTTACGAAGCTACTGTCTACCTTCCGATGCTATC 2575
 810 rGspValGluSerGlyProValValLeuLeuLysAsnAlaValSerTyr 826
 2576 GTAAGAATCCTGACTGCAGCAGCTCTCTTAATCAACAATACCTCGTGG 2625
 827 AspAlaProMetAlaAsnLeuAspSerArgAla..... 837
 2626 AAAACTACAGGAACGAATCTCTCAAGACAAGCTGTATCGGAAGACAGG 2675
 838TyrMetPheArgLeuThrAsnGlnArgAlaLeuHisArgLeuG 852
 2676 GATCTTTATGCTCTCT.....CCAAATCTTG 2704
 852 InThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSerTyr 868
 2705 AGGTCACAAGTAACCTATCTATGGAATTCGTGGATCTTCACGCAGCTAC 2754
 869 SerLeuAspLeuGlyThrThrTyrArgPhe 878
 2755 AATGCAGATCTTGGAGGTAAGTTCAGTTC 2784

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.X91990

seq_documentation_block:

ID X91990 standard; DNA; 1230025 BP.

XX AC X91990;

DT 13-SEP-1999 (first entry)

XX Nucleotide sequence of the complete genome of Chlamydia pneumoniae.

DE Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope; ss.

XX Chlamydia pneumoniae.

OS WO9927105-A2.

PD 03-JUN-1999.
 XX 20-NOV-1998; 98WO-1B01890.
 XX 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX (GEST) GENSET.
 XX Griffais R;
 PI WPI; 1999-357842/30.
 DR Genome sequence of Chlamydia pneumoniae
 XX
 PT Claim 1: Page 291-611; 1912pp; English.
 XX
 PS The present sequence represents the complete genome of Chlamydia
 CC pneumoniae, and encodes proteins Y34584-Y35879. C. pneumoniae causes
 CC respiratory disease such as pneumonia and bronchitis and is thought
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
 CC encoded by the open reading frames of the C. pneumoniae genome (see
 CC Y34584-Y35879) can be used in immunogenic compositions as vaccines.
 CC Vectors containing C. pneumoniae nucleotide sequences can also be
 CC used as immunogenic compositions, especially where the vector directs
 CC the expression of a neutralising epitope of C. pneumoniae.
 XX
 SQ Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

alignment_scores:

Quality: 930.00 Length: 910

Ratio: 1.745 Gaps: 33

Percent Similarity: 58.571 Percent Identity: 31.538

alignment_block:

US-09-677-752-4 x X91990

Align seg 1/1 to: X91990 from: 1 to: 1230025

85 ThrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerIleG1 101
 510813 ACAATGACAGATGGAATCTATCTATACTAATCTACAGGGGATGCTCAATCAC 510862
 101 nAsnPheArgPhe...LeuSerPheThrAspCysSerLysGluSerS 117
 510863 CAATCCAGGATCTCCGACAGCTCTAACCGCTTCTGCTTTAAAGAACTA 510912
 117 erProSerIleIleHisGlnLysAsnGly...GlnLeuSerLeuArgasn 132
 510913 CTGGGAATCTTTCTTTCAAGGCCACGGCTACCAATTTCTCTACAAAT 510962
 133AsnGlySe 135
 510963 ATCGATGCGGGAGCAACTGTACCTTTTACCAATACAGCTGCAATAAGCT 511012
 135 rMetSerPhe.....CysArgAsnH 142
 511013 TCTCTCTTTTCAGGATCTCTCTATTGTCCTACTAATAACACCAAGATG 511062
 142 IsAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSerLeuGln 158
 511063 CTACACAGGACACAGAGCCATCAAGTCCACAGAGCTTGTCTATTTCAG 511112
 159 HisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerLysGlyAsnG1 175
 511113 TCGAACTAT...AGTTGCTACTTTGGCCAAAACACTTTTCTATGACAATGG 511159
 175 yGlyAlaIleGlnAlaGlnThrPheSerLeuSerArgAsnValSerProI 192
 511160 AGGCGCCCTCCCAAGGCAGCTCTATCAGTCTATCGCTAAACCCCAAC...C 511206

```

512935 CAAG...GGAGAGAGCTCGTGGTATTAATCAACGCTTCGCGCTCGACATT 512981
729 CysAlaSerIleProIleValSer...AsnGlySerGlyLeuPheSerSe 744
:|||||:|||||:  :  :|||||:  :
512982 GCGAGCTCCCTACCAACACACTGCTTTAAGCCATGAGGCTCTCTCCAGC 513031
744 rPheSerIlePheSerIleGlyLeuGlyPheSerGlyThrGlnAspGlyP 761
|||||:|||||:  :  :|||||:  :
513032 GTATTTCTCTTCATCAAGATAGAAAGCTTCGTACATACACAGATAGCT 513081
761 heGluGlu...SerSerGlyGluIleArgSerPheSerAlaSerSerPhe 776
|||||:|||||:  :  :|||||:  :
513082 TCAAGAACGTAATACCTACCTGTCAGATCTTTCGATAGCGGTGATTTA 513131
777 ArgAsnIleSerLeuProIleGlyIleThrPheGluIleGlySerGlnLy 793
|||||:|||||:  :  :|||||:  :
513132 ATTAACGCTCTCTGCTGCTATTGGAATTACCTTCGAGAGATTCGAGAA 513181
793 sThrArgThrTyrTyrPheLeuGlyAlaTyrIleGlnAspLeuLysA 810
|||||:|||||:  :  :|||||:  :
513182 CGAGCGTGGCTCTTACGAAGCTACTGTCTACGTGCGGATGCTTATC 513231
810 rGaspValGluSerGlyProValValLeuLeuLysAsnAlaValSerTirp 826
||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
513232 GTAGAAATCTGACTGCACACAGCTCTCTTAATCAACAATACCTCTGG 513281
827 AspAlaProMetAlaAsnLeuAspSerArgAla..... 837
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
513282 AAAACTACAGGAAGCAATCTCTCAAGACAAGCTGGTTCGGAAGCAGG 513331
838 .....TyrMetPheArgLeuThrAsnGlnArgAlaLeuHisArgLeuG 852
|||||:|||||:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
513332 GATCTTTTATGCCTCTCT.....CCAAATCTTG 513360
852 InThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSerTyr 868
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
513361 AGGTCACAAGTAACCTATCTATGAAATTCGTGATCTTCACGCAGCTAC 513410
869 SerLeuAspLeuGlyThrThrArgPhe 878
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
513411 AATGCAGATCTTGGAGGTAAGTTCAGCTTC 513440

```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30853

seq_documentation_block:

ID A30853 standard; DNA; 3000 BP.

AC A30853;

DT 29-AUG-2000 (first entry)

DE Chlamydia antigen CPN100639 full length coding sequence.

KW Chlamydia antigen; diagnosis; infection; community acquired pneumonia;
 KW therapy; upper respiratory tract disease; bronchitis; sinusitis;
 KW asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma;
 ds.

XX Chlamydia pneumoniae.

XX Key Location/Qualifiers
 FT CDS 101..2893
 FT /*tag= a
 FT /product= Chlamydia antigen CPN100639

XX W0200032794-A2.

PN 08-JUN-2000.

XX 01-DEC-1999; 99WO-CA01147.

XX 01-DEC-1998; 98US-0110339.

PR 01-DEC-1998; 98US-0110340.

```

PR 01-DEC-1998; 98US-0110427.
PR 01-DEC-1998; 98US-0110428.
XX 01-DEC-1998; 98US-0110438.
XX (CONN-) CONNAUGHT LAB LTD.
XX Murdin AD, Oomen RP, Wang J;
XX WPI: 2000-412339/35.
XX P-PSDB: Y90240.
XX Nucleic acids encoding polypeptide antigens from Chlamydia useful for
PT preventing, diagnosing and treating diseases such as community acquired
PT pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset
PT asthma -
XX
PS Claim 2: Fig 7: 174pp: English.
XX
CC This sequence encodes a Chlamydia antigen of the invention, designated
CC CPN100639. The nucleic acids (and their complementary sequences) may be
CC used as diagnostic agents for detecting the presence of nucleic acids
CC encoding Chlamydia antigens in samples according to standard methods,
CC and therefore, for diagnosing Chlamydia infections. For example, they may
CC be used as primers and probes for diagnostic polymerase chain reaction
CC (PCR) assays. Antisense sequences may be used to down regulate
CC expression of the proteins and may be used to treat infections. The
CC nucleic acids may also be used to produce the protein antigens they
CC encode according to standard recombinant DNA methodologies. The
CC proteins may then be used as antigens for the production of antibodies
CC (i.e. as vaccines) for preventing infection by Chlamydia. The
CC antibodies may also be used as diagnostic reagents for detecting
CC infections. Chlamydia is a pathogen implicated in the development of
CC (for example) community acquired pneumonia, upper respiratory tract
CC disease (especially bronchitis and sinusitis, asthmatic bronchitis,
CC adult-onset asthma and acute exacerbations of asthma in adults.
XX
SQ Sequence 3000 BP; 858 A; 726 C; 592 G; 824 T; 0 other:

```

alignment_scores:

Quality: 912.50 Length: 840
 Ratio: 1.702 Gaps: 26
 Percent Similarity: 63.810 Percent Identity: 30.833

alignment_block:

US-09-677-752-4 X A30853

Align seg 1/1 to: A30853 from: 1 to: 3000

```

98 LeuSerIleGlnAsnPheArgPheLeuSerPhe.....Th 109
|||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
455 CTAATCTTCACAGGATTTCTTAACCTTTCTTCTTCTTCAGCTCTCGGAAC 504

109 rAspCysSerSerIleGluSerSerProSerIleIleHisGlnLysAsnG 126
|  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
505 TACAGTTGCTTCAGGAAAAAGTACTTTAAGT.....TCTGCAG 542

126 IyGlnLeuSerLeuArgAsnAsnGlySerMetSerPheCysArg..... 140
||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
543 GAGCCTTAATCTTACCGATAATGGAACGATCTCTTTAGCCAAACGTC 592

141 ...AsnHisAlaGluGlySerGlyAlaIleSerAlaAspAlaPheSe 156
|||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
593 TCCAATGAAGCTAATAACAATGGCGGAGCGATCACCACAAAACCTCTTC 642

156 rLeuGlnHisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerLysG 173
|  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
643 TATTCTGGGAATACCTCTTCTTATAACCTTCACCTAGTAATAAGCCAAAA 692

173 IyAsnGlyGlyAlaIle...GlnAlaGlnThrPheSerLeuSerArgAsn 188
|||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
693 AATTAGGTGGAGCGATCTATAGCTTCGCGCTGCAAGTATTTCAGGAAC 742

```


806 nAspLeuLysArgAspValGlnSerGlyProValValLeuLeuLysAsnA 823
 2571 TGATGTGTTATCGTAAATCCCGCTTCGCCTACTCTCTTAATGTGTCAGTG 2620
 823 laValSerTrpAspAlaProMetAlaAsnLeuLeuAspSerArgAlaTrMet 839
 2621 GACCTCTTGGACTTCGCTATGTAATAAACCCTCCACGACAGCCCTCTTA 2670
 840 PheArgLeuThrAsnGlnArgAlaLeuHis...ArgLeuGlnThrLeuLe 855
 2671 GCAAGTGCTGGAAGCCATCTGACTCTCTCCCTCATGTAGAACTCTCTGG 2720
 855 uAsnValSerCysValLeuArgGlyGlnSerHisSerTyrSerLeuAspL 872
 2721 GGAAGCTGCTTATGAGCTTGTGCTGCTGACACACATCTACAATGTAGATT 2770
 872 euGlyThrThrArgPhe 878
 2771 GTGGGCTAGACTATCTATTC 2790

seq_name: /SID52/gcgdata/geneseg/geneseqn/NA1999.DAT:X06817

seq_documentation_block:
 ID: X06817 standard; DNA; 2815 BP.

XX X06817;

XX AC XX

XX DT 26-APR-1999 (first entry)

XX DE Chlamydia pneumoniae surface exposed protein Omp5 DNA.

XX KW Omp5; outer membrane protein 5; surface exposed protein; antigen;

XX KW infection; diagnosis; vaccine; atherosclerosis; asthma; ss.

XX OS Chlamydia pneumoniae.

XX FH Key Location/Qualifiers

XX FT CDS 1..2787

XX FT /*tag= a

XX PN W09858953-A2.

XX PD 30-DEC-1998.

XX PF 19-JUN-1998; 98WO-DK00266.

XX PR 23-JUN-1997; 97DK-0000744.

XX PA (BIRK/) BIRKELUND S.

XX PA (CHR1/) CHRISTIANSEN G.

XX PI Birkelund S, Christiansen G, Knudsen K, Madsen A;

XX PI Mygind P;

XX XX WPI: 1999-105610/09.

XX DR P-PSDB; W88418.

XX XX Species-specific test for identifying mammals infected with

XX PT Chlamydia pneumoniae - comprises detecting antibodies specific for

XX PT outer membrane proteins of C. pneumoniae or nucleic acids encoding

XX PT these proteins

XX PS Claim 6; Page 42-43; 115pp; English.

XX CC This DNA sequence codes for the novel 97.2 kDa surface exposed

CC the other, and encode polypeptides of about 89.6-100.3 kDa and 15 in
 CC about 56.1 kDa. The invention provides a new species specific test
 CC for identifying mammals (including humans) infected with Chlamydia
 CC pneumoniae. The test comprises detecting antibodies specific for
 CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
 CC membrane proteins, especially by PCR. The proteins are also used
 CC in the diagnosis of C. pneumoniae infection in mammals. The
 CC nucleic acids and proteins can also be used in the immunization of
 CC mammals, the nucleic acids being particularly useful as DNA
 CC vaccines for effecting in vivo expression of antigens. The
 CC vaccines may also prevent atherosclerosis and bronchial asthma,
 CC which are possibly associated with C. pneumoniae.

XX Sequence 2815 BP; 790 A; 575 C; 626 G; 824 T; 0 other;

alignment_scores:
 Quality: 883.00 Length: 942
 Ratio: 1.541 Gaps: 31
 Percent Similarity: 60.828 Percent Identity: 30.573

alignment_block:

US-09-677-752-4 x X06817 ..

Align seg 1/1 to: X06817 from: 1 to: 2815

12 CysAlaAlaLeuSerSerThrAlaValLeuPheGlyGlnAspProle 28
 58 TGTTCACCTGTTTTGCTGCAACTGCTGAAATATATAGGCCCTCTCATAG 107
 28 u.....GlyGluThrAlaLeuLeuThrLysAsnProAsnHisValAlC 43
 108 CTTTGTGCGGAAGTACTAACACAGGCACCTATATCTCTCTCTCTCTCTCTCT 157
 43 ys.....ThrPhePheGluAspCysThrMetGluSerLeu... 54
 158 CTGGATAGACTATCTCTGACAGGAGATATACTCTGCAAAACCTTGGG 207
 55 ...PheProAlaLeuCysAlaHisAlaSerGlnAsp.....AspProle 68
 208 GATTTCGGCAGCTTAAACGAAGGGTCTTTTCTGACACTACGGAATCTTT 257
 68 uTyrValLeuGlyAsnSerTyrCysTrpPheValSerLysLeuHisIleT 85
 258 AGCTTTGCGCGTAAAGGGTACTCA.....CTTCTTTTAAATATTA 301
 85 hrAspProLysGluAlaLeuPheLysGluLysAspLeuSerIleGln 101
 302 AGTCTAGTCTGAGGC.....GCAGCACTTCTGTGTACA 336
 102 AsnPheArgPheLeuSerPheThrAspCysSerSerLysGlu.....Se 116
 337 ACTGATAAAATCTGCTGCTAACAGGATTTTCGAGTCTTACTTTCTTAGC 386
 116 rSerProSerIleIle.....HisGlnL 124
 387 GGCCCATCATCGTAAATCAACCCCTCAGGAAAGGTGTCAGTTAAAT 436
 124 ysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPheCysArg 140
 437 GTGAGGGGATCTTACATTTGATAACAATGGAACATATTTATTAAACAA 486
 141 AsnHisAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSerle 157
 487 GATTACTGTGAGGAAATGCGGAGCCATTCTTACCAAGAAATCTTTCTTT 536
 157 uGlnHisAsnTyrLeuPheThrAlaPheGluAsnSerSerLysGly. 173
 537 GAAAAACAGCACCGGATTCGATTTCTTTTGAAGGAATAAATTCGAGCGCAA 586
 174AsnGlyGlyAlaIleGlnAlaGln...ThrPheSerLeuSer 186

CC cassette of (3) as part of an extrachromosomal element or integrated into
CC the genome of a host cell as a result of induction of the expression
CC cassette into the host cell, and the cellular progeny of the host cell;
CC (5) a method for producing a Pl comprising growing a cell of (4) where
CC the protein is expressed and isolating the protein free of other
CC proteins; (6) a purified polypeptide composition comprising at least 50
CC weight % of Pl; and (7) a monoclonal antibody binding specifically to the
CC peptide of (6).
XX
XX
SQ Sequence 273254 BP: 76423 A: 51054 C: 61965 G: 83812 T: 0 other:

Sequence 273254 BP: 76423 A: 51054 C: 61965 G: 83812 T: 0 other;
XX

```

alignment_scores:
  Quality: 866.00      Length: 942
  Ratio: 1.509        Gaps: 31
Percent Similarity: 60.934  Percent Identity: 30.573

alignment_block:
  US-09-677-752-4 x C81914  ..

Align seg 1/1 to: C81914 from: 1 to: 273254

12  CysAlaAlaIleLeuSerSerThrAlaValLeuPheGlyGlnAspProLe 28
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98251  TGTTCACACTGTTTTGCTGCAACTGCTCAAAATATAGGCCCTCTGATAG 98300
28  u.....GlyGluThrAlaLeuLeuThrLysAsnProAsnHisValValC 43
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98301  CTTTCACGGAAGTACTAACACAGGACCTATACTCTCTAAATAACGACTA 98350
43  ys.....ThrPheGluAspCysThrMetGluSerLeu... 54
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98351  CTGGAATAGACTATACTCTGACAGGAGATATAACTCTGCAAAACCTTGGG 98400
55  ...PheProAlaLeuCysAlaHisAlaSerGlnasp.....AspProLe 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98401  GATTCGGCAGTTTAAACGAAGGTGTTTTCGTGACACTACGGAAATCTTTT 98450
68  uTyrValLeuGlyAsnSerTyrCysTrpPheValSerLysLeuHisIleT 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98451  AGCTTTGCGGTAGGGGTACTCA.....CTTCTCTTTTAAATATTA 98494
85  hrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerLleGln 101
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98495  AGTCTAGTCTGAAGGC.....GCAGCACCTTCTGTTTACA 98529
102  AsnPheArgPheLeuSerPheThrAspCysSerSerLysGlu.....Se 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98530  ACTGATATAAAATCTGTCGCTAACAGAGTTTTCGAGTCTTACTTCTTAGC 98579
116  rSerProSerIleIle.....HisGlnL 124
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98580  GGCCCCATCATCGGTAAATCACACCCCTCAGGAAAAGTGCAGTTAAAT 98629
124  ysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPheCysArg 140
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98630  GTGGAGGGATCTTACATTGTATAACAATGGAACACTTTTATTATTAACAA 98679
141  AsnHisAlaGluClySerGlyGlyAlaIleSerAlaAspAlaPheSerLe 157
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98680  GATTACTGTGAGAAATCGGGAGGCATTTCTACCAAGAAATCTTTCITTT 98729
157  uGlnHisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerLysGly 173
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98730  GAAAACAGCACCGGGATCGATTCTTTTCAGGGGATAAATCGAGCGCAA 98779
174  .....AsnGlyGlyAlaIleGlnAlaGln...ThrPheSerLeuSer 186
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98780  CAGGAAAAAAGGTGGGGCTATTGTGCTACTGTTACTGTAGATATTACA 98829
187  ArgAsnValSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnGl 203
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98830  AATAATACGGCTCCTACCCCTCTTCTCGAACAAATATTGCTGAAGCTGCAGG 98879

```


seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:Z61508

seq_documentation_block:

ID Z61508 standard: DNA; 3050 BP.

XX

AC Z61508;

XX

DT 19-JUN-2000 (first entry)

XX

DE DNA encoding the CPN100394 polypeptide.

XX

KW CPN100394; Chlamydia infection; immune response; vaccine; ss.

XX

OS Chlamydia pneumoniae.

XX

FH Key Location/Qualifiers

FT CDS 101..2938

FT /*tag= a

FT sig_peptide 101..226

FT /*tag= b

XX

PN WO200011183-A2.

XX

PD 02-MAR-2000.

XX

PF 18-AUG-1999; 99MO-IB01449.

XX

PR 20-AUG-1998; 98US-0097187.

XX

PR 20-AUG-1998; 98US-0097188.

XX

PR 20-AUG-1998; 98US-0097189.

XX

PR 20-AUG-1998; 98US-0097190.

XX

PR 20-AUG-1998; 98US-0097195.

XX

PR 20-AUG-1998; 98US-0097196.

XX

PR 20-AUG-1998; 98US-0097197.

XX

PR 27-AUG-1998; 98US-0097191.

XX

PR 17-AUG-1999; 99US-0376770.

XX

PA (CONN-) CONNAUGHT LAB LTD.

XX

PI Murdin AD, Oomen RP;

XX

XX WPI: 2000-224703/19.

XX

DR P-PSDB; Y69368.

XX

XX Novel antigens and corresponding DNA molecules that can be used to

PT prevent, treat and diagnose disease caused by Chlamydia infection in

PT mammals, especially humans.

XX

NS Claim 1; Fig 13A-F; 201pp: English.

XX

CC Z61502-09 encode Chlamydia pneumoniae polypeptides. The polypeptides are present in the bacterial membrane structure, in the external vicinity of the membrane structure, in the inclusion membrane structure, in the external vicinity of the inclusion membrane structure, and in the cytoplasm of the infected cell. The polypeptides may be used to prevent, treat and detect the presence of Chlamydia infection and/or the presence of Chlamydia in a sample. The polypeptides may also be used to induce an immune response in a mammal. The vaccine vector comprising the polynucleotides is used to induce an immune response in a mammal. Antibodies directed against the polypeptides may also be used therapeutically to treat and/or prevent a Chlamydia infection.

XX

SQ Sequence 3050 BP; 897 A; 622 C; 695 G; 836 T; 0 other;

alignment_scores:

Quality: 851.00

Ratio: 1.567

Percent Similarity: 62.702

Percent Identity: 28.637

alignment_block:

US-09-677-752-4 x Z61508

Align seg 1/1 to: Z61508 from: 1 to: 3050

```

74 SerTyrCysTrpPheValSerLysLeuHisIleThrAspProLysGluAl 90
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
449 AGCTTCACAGATTTCCTTCTCTGGTGATCACAGAAATCTCCAATACGC 498
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
90 aLeuPheLysGluLysGlyAspLeu.....SerIleG 101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
499 TGTTACTACAGGAAAGTAGCTAGTCAGTTAGTGCAGTCCAACCTGC 548
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 InAsnPheArgPheLeuSerPheThrAspCysSerSerLysGluSerSer 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
549 AAGATATAACACTCTAGTTCTTACAAGCAATGCCTCTCTCGAAGATGGT 598
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
118 ProSerIleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsnG1 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
599 GCCGTGATT.....AAGGAAACTCTCTGTGATTTCAGGAATCAAAAA 642
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
134 ySerMetSerPheCysArgAsnHisAlaGluGlySerGlyGlyAlaIleS 151
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
643 TAGTGCAGATTTTGGACAAATACATCTTCGAAAAAAGAGGGCGCATCT 692
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151 er...AlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
693 CCAGGACTCAAGGACTTACCATAGAGATAACTTAGGGACGCTAAAGTTC 742
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
167 GluGluAsnSerSerLysGlyAsnGlyGlyAlaIleGln.....Al 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
743 AATGAAAAACAAAGCAGTGACCTCAGGAGCGCCTTAGATTAGGACCGCG 792
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
180 acLntThrPheSerLeuSerArgAsnValSerProIleSerPheAlaArgA 197
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
793 GTCTACATTTCACGCAACCATGAG.....TTGATATTTTCACAAA 833
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
197 snArgAlaAsp.....LeuAsnGlyGlyAlaIleCysCysSer... 209
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
834 ATAAGACTTCTGGGAATGCTGCAATGCGGAGGCATAAATGTCTCAGGG 883
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
210 AsnLeuIleCysSerGlyAsnValAsnProLeuPheThrGlyAsnSe 226
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
884 GACCTTACATTTACTGATAACACTTCTTTGTTACTTCAAGAAAAATAGCAC 933
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
226 rAlaThrAsnGlyGlyAlaIleCysCysIleSerAspLeuAsnThrSerG 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
934 AATCAGGATGCTGGAGCTTTGTGT.....A 959
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
243 LuLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAlaSer 259
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
960 GCACAGGAAACCATAGCATTACCGGTAGTATTCTATCAATGTGATAGGA 1009
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
260 AsnSerAlaLysGluLysGlyAlaIleTyrAlaLysHisMet...Va 275
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1010 AATACTTCAGGACAAAAAGGAGGAGCGATTCTCGACGCTTCTCTCAAGAT 1059
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
275 lLeuArgTyrAsnGlyProValSerPheIleAsnAsn.....S 288
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1060 TTTGGGAGGGCAGGAGCGCGCTCTCTTTCTTAATACCTAGTACTCATG 1109
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
288 erAlaLysIlleGlyAlaIleAlaIleGlnSerGlyGlySerLeuSer 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1110 CCAGCCCTCTAGAGGTGCCATTTTATCAACACAGGAGGATCCTTGCAG 1159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
305 lLeuAlaGlyGluGlySerValLeuPheGlnAsnAsnSerGlnArgTh 321
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1160 CTCCTTCACTCAAGGAGGGGATATCGTATTTCGAGGGGAATCAGGTCACTAC 1209
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
321 rSerAspGlnGlyLeuVal.....ArgAsnAlaIleTyrLeuGluLysA 336
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1210 AACAGCTCCAATGCTACCACCTAAGAGAAATGTAATTCACCTCGAGACA 1259
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
336 spAlaIleLeuSerSerLeuGluAlaArgAsnGlyAsp...IleLeuPhe 351

```

```

DT XX 29-AUG-2000 (first entry)
DE XX Chlamydia antigen CPN100634 full length coding sequence.
XX
KW Chlamydia antigen; diagnosis: infection; community acquired pneumonia;
KW therapy: upper respiratory tract disease; bronchitis; sinusitis;
KW asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma;
KW ds.
XX
OS Chlamydia pneumoniae.
XX
XX Key Location/Qualifiers
XX CDS 101..2887
XX /tag= a
XX /product= Chlamydia antigen CPN100634
XX
XX WO200032794-A2.
XX
XX 08-JUN-2000.
XX
XX 01-DEC-1999; 99WO-CA01147.
XX
XX 01-DEC-1998; 98US-0110339.
XX 01-DEC-1998; 98US-0110340.
XX 01-DEC-1998; 98US-0110427.
XX 01-DEC-1998; 98US-0110428.
XX 01-DEC-1998; 98US-0110438.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Murdin AD, Oomen RP, Wang J;
XX
XX WPI; 2000-412339/35.
XX P-PSDB; Y90236.
XX
XX Nucleic acids encoding polypeptide antigens from Chlamydia useful for
XX preventing, diagnosing and treating diseases such as community acquired
XX pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset
XX asthma.
XX
XX Claim 2; Fig 1; 174pp; English.
XX
XX This sequence encodes a Chlamydia antigen of the invention, designated
XX CPN100634. The nucleic acids (and their complementary sequences) may be
XX used as diagnostic agents for detecting the presence of nucleic acids
XX encoding Chlamydia antigens in samples according to standard methods,
XX and therefore, for diagnosing Chlamydia infections. For example, they may
XX be used as primers and probes for diagnostic polymerase chain reaction
XX (PCR) assays. Antisense sequences may be used to down regulate
XX expression of the proteins and may be used to treat infections. The
XX nucleic acids may also be used to produce the protein antigens they
XX encode according to standard recombinant DNA methodologies. The
XX proteins may then be used as antigens for the production of antibodies
XX (i.e. as vaccines) for preventing infection by Chlamydia. The
XX antibodies may also be used as diagnostic reagents for detecting
XX infections. Chlamydia is a pathogen implicated in the development of
XX (for example) community acquired pneumonia, upper respiratory tract
XX disease (especially bronchitis and sinusitis, asthmatic bronchitis,
XX adult-onset asthma and acute exacerbations of asthma in adults.
XX
XX Sequence 2950 BP; 851 A; 670 C; 596 G; 833 T; 0 other;
XX
alignment_scores:
XX Quality: 849.50 Length: 959
XX Ratio: 1.467 Gaps: 32
XX Percent Similarity: 60.375 Percent Identity: 27.737
XX
alignment_block:
XX US-09-677-752-4 x A30847 ..
XX
Align seg 1/1 to: A30847 from: 1 to: 2950

```

CC nucleic acids and proteins can also be used in the immunization of
 CC mammals, the nucleic acids being particularly useful as DNA
 CC vaccines for effecting in vivo expression of antigens. The
 CC vaccines may also prevent atherosclerosis and bronchial asthma,
 CC which are possibly associated with C. pneumoniae.
 XX
 SQ Sequence 3200 BP; 946 A; 708 C; 623 G; 923 T; 0 other;

alignment_scores:
 Quality: 849.50 Length: 959
 Ratio: 1.467 Gaps: 32
 Percent Similarity: 60.375 Percent Identity: 27.737

alignment_block:

US-09-677-752-4 x X06816

Align seg 1/1 to: X06816 from: 1 to: 3200

```

      7 AsnPheCysCysLeuGlyGlnThrAlaLeuSerThrAlaValLeuPh 23
      ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
    284 AACTTTTATCAGCTGATGATGCTTTTAAATGGAATATCC..... 322
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    23 eGlyGlnAspProLeuGlyGlnThrAlaLeuSerThrAlaValLeuPh 40
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    323 .....ATTGAGGAACTTTTACTCCAAACTTCAGGCAAC 359
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    40 isValValCysThrPhePheGluAspCysThr...MetGluSerLeuPhe 55
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    360 ATATTCTTAACAGGAGATGCTTTTACGAGCCTGGAAGGCACTC 409
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    56 proAlaLeuGlyGlnAlaSerGlnAspProLeu.TyrValLeuG 72
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    410 CCTTATCTGACAGTTGTTTAAAGCAACACCGGCAATCTTACCTTCTG 459
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    72 LysAsnSer.....TyrCysTrpPheValSer 80
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    460 GGAAGGCTCATAGCTTAACGTTTGGCTTTATAGATGCTGGCACTCATGC 509
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    81 LysLeuHisIleThrAspProGlyGlnAlaLeuPheGlyGlyA 97
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    510 AGGTGCTGC.....TGCATCTACAACAGCAATAAGA 541
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    97 spLeuSerIleGlnAsnPheArgPheLeuSerPheThrAspCysSer 113
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    542 ATCTTACCTTCTCAGGTTTCTTCTTACTGAGTTT..... 576
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    114 LysGluSerSerProSer...IleIleHisGlnLysAsnGlyGlnLeu 129
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    577 ...GATTCCTCTCTAGCACACGCTTACTACAGGTCAGGGAACGCTTTC 623
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    129 r.....LeuArgAsnAsnGlySerMetSerPheCysA 140
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    624 CTCAGCAGGAGCGTAAATTTAGAAAATATTCGTAAACTTGTAGTTGCTG 673
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    140 rGAsnHisAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPhe 156
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    674 GGAATTTTCTACTGAGATGCTGAGCTTATCAAGGAGCGCTTCTTCTCT 723
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    157 LeuGlnHisAsnTyrLeuPheThrAlaPheGluGluAsnSerLysG 173
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    724 TTAACGTCACCTCTGGAGATGCTCTTTTACTAACAACCTTTCATCAAC 773
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    173 rAsnGlyGlyAlaIleGlnAlaGlnThrPheSer...LeuSerArgAsnV 189
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    774 NAAGGAGGAGCAATTGCTACTACAGCGCGCTCCGATAGCAATAACA 823
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    189 aLysProIleSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAla 205
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    824 CAGGTTATGTTAGATTCTCTTATCAATACAGCTTCTAGCTCAGGAGCGCT 873
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    206 IleCysCysSerAsnLeuIleCysSerGlyAsnValAsnProLeuPhe 222
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  
```

```

846 ArgAlaLeuHisArgLeuGlnThrLeuLeuAsnValSerCys..... 859
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2780 .....AACTAGCTTACAACCTCCAAATTTGAGCTCTT 2811
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
860 .....ValLeuArgGlyGlnSerHisSerTyrSerLeuA 871
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2812 CGGACATTACGCTATGGAACCTCCGTGATCTTCAAGGAACACTACAATGTAG 2861
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

871 spLeuGlyThrThrTyrArgPhe 878

2862 ATGTTGTTACCAAACTCCGATTC 2884

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.X06816

seq_documentation_block:

ID X06816 standard; DNA; 3200 BP.

XX X06816;

XX 26-APR-1999 (first entry)

XX Chlamydia pneumoniae surface exposed protein Omp4 DNA.

XX Omp4; outer membrane protein 4; surface exposed protein; antigen;
 XX infection; diagnosis; vaccine; atherosclerosis; asthma; ss.

XX Chlamydia pneumoniae.

XX Key Location/Qualifiers
 XX CDS 205..2991
 XX FT /*tag= a

XX W0958953-A2.

XX 30-DEC-1998.

XX 19-JUN-1998; 98WO-DK00266.

XX 23-JUN-1997; 97DK-0000744.

XX (BIRK/) BIRKELUND S.

XX (CHRI/) CHRISTIANSEN G.

XX Birkelund S, Christiansen G, Knudsen K, Madsen A;

XX Mygind P;

XX WPI; 1999-105610/09.

XX P-PSDB; W88417.

XX Species-specific test for identifying mammals infected with

XX Chlamydia pneumoniae - comprises detecting antibodies specific for
 XX outer membrane proteins of C. pneumoniae or nucleic acids encoding
 XX these proteins

XX Claim 6; Page 35-40; 115pp; English.

XX This DNA sequence codes for the novel 98.9 kDa surface exposed

XX protein Omp4 (see W88417) of the human respiratory pathogen

XX Chlamydia pneumoniae. By generating antibodies against C.

XX pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)

XX was obtained which reacted with outer membrane proteins. The

XX antibody was used to identify the genes (see X06816-27) encoding

XX Omp4-Omp15 proteins (see W88417-28) in an expression library of

XX C. pneumoniae DNA. The genes are situated in 2 gene clusters:

XX Omp12, 11, 10, 5, 4, 13 and 14 in one cluster and Omp6, 7, 8, 9 and 15 in

XX the other, and encode polypeptides of about 89, 6-100.3 kDa and

XX about 56.1 kDa. The invention provides a new species specific test

XX for identifying mammals (including humans) infected with Chlamydia

XX pneumoniae. The test comprises detecting antibodies specific for

XX Omp4-Omp15 or detecting nucleic acid fragments encoding these outer

XX membrane proteins, especially by PCR. The proteins are also used

XX in the diagnosis of C. pneumoniae infection in mammals. The


```

2494 GGAGGTGAGGTCCTCACTTACTTAGCGGAGATCTTTCAATTTAGCAGT 2543
781 uProfileGlyThrPheGluLysLysSerGlnLysThrArgPheTyrT 798
2544 TCCTATTGGCGTGAAGTTTGAGAGATTTTCAGACTGTAAAGGGGATCTT 2593
798 yTyrPheLeuGlyAlaTyrIleGlnAsp...LeuLysArgAspValGlu 813
2594 ATGAACATTACCTTGCCTTATGTTCTCTGATGTGATTCGCAAGATCCCAAG 2643
814 SerGlyProValValLeuLeuLysAsnAlaValSerTrpAspAlaPhe 830
2644 AGC...ACGGCAACATTGGCTAGTGGAGCTACGTGGAGCACCCACGG 2687
830 tAlaAsnLeuAspSerArgAlaTyrMetPheArgLeuThrAsnGlnArg 847
2688 AACCAATCTCTCCAGACAAGGATTACAACTGCGTTTAGGGACCACTGTC 2737
847 lAlaLeuHis...ArgLeuGlnThrLeuLeuAsnValSerCysValLeuArg 862
2738 TCATAAATCCTGGAATTCAGGTGTTCAGTCACGGAGCTATTGAATTGCGG 2787
863 GlyGlnSerHisSerTyrSerLeuAspLeuGlyThrThrTyrArgPhe 878
2788 GGATCCTCTCGTAATTATAACATCAATCTCGGGGGTAATACCGGATTT 2835
seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.A27342

```

seq_documentation_block:

```

ID_A27342 standard; DNA; 3150 BP.
XX
XX AC A27342;
XX DT
XX 10-AUG-2000 (first entry)
XX DE Chlamydia POMP91B precursor gene.
XX
XX Chlamydia; POMP91B; respiratory tract disease; infection;
KW bronchitis; sinusitis; pneumonia; atherosclerosis; asthma; vaccine;
KW immunogen; ds.
XX
XX Chlamydia pneumoniae.
XX
XX Key Location/Qualifiers
XX CDS 101..3022
XX FT /*tag= a
XX FT /product= "POMP91B precursor"
XX
XX WO200026239-A2.
XX
XX 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-GB03622.
XX
XX 02-NOV-1998; 98US-0106590.
XX 07-MAY-1999; 99US-0133071.
XX 29-OCT-1999; 99US-0430723.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Murdin AD, Oomen RP, Dunn PL;
XX
XX WPI; 2000-365571/31.
XX P-PSDB; Y96274.
XX
XX Novel Chlamydia POMP91B precursor protein antigen, used for vaccination
XX and protection against Chlamydia infection
XX
XX Claim 1; Fig 1; 97pp; English.
XX
XX The present sequence is the Chlamydia pneumoniae, POMP91B precursor
XX protein gene. Infection by Chlamydia can result in respiratory tract

```

CC diseases such as bronchitis, sinusitis and pneumonia, asthma and
 CC atherosclerosis. The gene, protein and antibodies can be used as
 CC immunogens to induce an immune reaction in humans which has the effect
 CC of vaccinating the person. They can also be used to diagnose and treat
 CC those infected with the parasite.
 XX
 SQ Sequence 3150 BP: 886 A: 760 C: 622 G: 882 T: 0 other:

alignment_scores:

Quality: 811.00 Length: 954
 Ratio: 1.438 Gaps: 28
 Percent Similarity: 59.119 Percent Identity: 26.520

alignment_block:

US-09-677-752-4 x A27342

Align seq 1/1 to: A27342 from: 1 to: 3150

```

37 AsnProAsnHisValValCysThrPheGluAspCysThrMetGluSe 53
245 GATCCTAGAGGGACACCTCTATTATTTTCAGGGGATCTCTACATTGCGAA 294
53 rLeuPheProAlaLeuCysAlaHisAlaSer.....GlnA 65
295 TCTTGATAATGCCATATCCAGAACCTCTCCAGTTGCTTTAGCAATAGGG 344
65 sPaspProLeuTyrValLeuGlyAsnSer.....TyrCysTrpPhe.Va 79
345 CGGGAGCACTACAAATCTTAGGAAAGGTGGGTTTCTCTCTTTAAAT 394
79 lSerLysLeuHisIleThrAspProLys..GluAlaLeuPheLysGluLy 95
395 ATCCGTTCTTCAGCTGAGCGAGCGCGATGATTAGTGTAAATCACCCAAA 444
95 sGlyAsp.....LeuSerIleGlnAsnPheArgPheLeuSerPheT 109
445 TCCTGAACTATGTCCTTGAGTTTTCAGGATTTAGTCAGATGATCTTCG 494
109 hrAspCysSerSerLysLysLysSerProSer.....Ileile 121
495 ATACTGTGATCTTTGACTTCAGATACCTCAGCGAGTAAATGTCATACCT 544
122 HisGlnLys.....AsnGlyGlnLeuSerLeuArgAsnAsnGl 134
545 CACGCATCGCGGATTTACGCTACAAACGCCCATGCTTTACAAACAATGA 594
134 ySerMetSerPheCysArgAsnHisAlaGluGlySerGlyGlyAlaIles 151
595 CTCATATACTATCCAAATACAAACCGTTCTGCGAGGATTTGGAGCTGCCATTC 644
151 erAlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPheGlu 167
645 GAGCACAAGCATCACAAATAGAAATACGAAAGAGAGCCTTCTCTTTAAT 694
168 GluAsnSerSerLysGlyAsnGlyGlyAlaIleGln...AlaGlnThrPh 183
695 GGTAATCGATCCATCTCTAATGGAGGGGCCCTCACGGGATCTGCGAGCAT 744
183 eSerLeuSerArgAsnValSerProIleSerPheAlaArgAsnArgAla 200
745 CACCTCATCAACAATAGCGCTCTCTGTGATTTTCTCAACGAATGCTACAG 794
200 spLeuAsnGlyGlyAlaIleCysCysSer.....AsnLeuIleCysSer 214
795 GGATCTATGGTGGGCTATTTACCTTACCGGAGGATCTATGCTCACCTCT 844
215 GlyAsnValAsnProLeuPhePheThrGlyAsnSerAlaThrAsnGlyGl 231
845 GGGACCTCTCAGGAGTCTTGTTCGTTAATAATAGCTCGCGCTCAGGAGG 894
231 yAlaIleCysCysIleSerAspLeuAsnThrSerGluLysGly..... 245

```



```

681 .....GlnAlaG 683
2118 GCACCTCATACCTACAAACCCAAAACATATTTGAGATTAGGAAGACAA 2167
683 lnLeuAlaThrSerLeuMetLys ..... 690
2168 AGTTTCTGAATCAGCTATAGAAAAATTCCTAGCGAAATTCCTAGCC 2217
691 IleSerAlaGlnAlaCysTyrAsnGlnSerIleHisGluLeuLysThrLy 707
2218 TTGCATGTCCAAGTTTCGTTCCAGCCATTTCAGACAACCGTATGGAAACGCA 2267
707 sTyrArgSerPheSerLysGluGlyPheGlySerTrpHisSerValAlav 724
2268 CTATACCTCATTTG...CCAGAATCCGAAGGTTCTTGGAGCAACGAGTGT 2314
724 alSerGlyGluValCysAlaSerIlePro...IleValSerAsnGlySer 739
2315 TAGCTGGGTGATCGGCTAGACCTTCCTTTGTTCTTTCCACCCACAT 2364
740 GlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGlyPheSerGl 756
2365 CCTCTTTTCAGACCTTCATTCCACAGATGAAGTCGAATGGTTTATGT 2414
756 yThrGlnAspGlyPheGluGluSerSerGlyGluIleArgSerPheSera 773
2415 ATCACAATAAGCTTCTTCGAAGCTCTAGTGTGGCGGCTGTTTAGTA 2464
773 laSerSerPheArgAsnIleSerLeuProIleGlyIleThrPheGluLys 789
2465 TTGGAAGGCTGTTAACTCTCGATTCCTGTGGTGCGAAATTCCTGCAG 2514
790 LysSerGlnLysThrArgThrTyrTyrPheLeuGlyAlaTyrIleGl 806
2515 GGGATATCGGAGATTCCTACACCTATCATCTCAGGATCTTTGTTC 2564
806 nAspLeuLysArgAspValGluSerGlyProvalValLeuLeuLysAsnA 823
2565 CGATGCTATCGTAACAATCCCAATCTACAGCGACTCTTGTGATGAGCC 2614
823 laValSerTrpAspAlaProMetAlaAsnLeuAspSerArgAlaTyrMet 839
2615 CAGACTCTTGGAAATTCGGGTGCAATCTTCAAGACAGCCATTTT 2664
840 PheArgLeuThrAsnGlnArgAlaLeuHisArgLeuGlnThrLeuLeuAs 856
2665 CTGAGGGGTAGCAAC.....AACTAGCTCTACAA 2693
856 nValSerCys.....ValLeuArgGlyGlnS 865
2694 CTCCAATTGTGAGCTCTTCGGACATTAGCCTATGGAACCTCCGTGATCTT 2743
865 erHisSerTyrSerLeuAspLeuGlyThrThrTyrArgPhe 878
2744 CAAGGAACACTAGATGTGTGGTACCACAACTCCGATTC 2784

```

OM of: US-09-677-752-4 to: GenEmbl.* out_format : pfs
 Date: May 6, 2001 8:33 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL-frame*_p2n.model -DEV=xlp
 -Q/cgnt2/_USPTO.spool/US09677752/runat_04052001_132443_4919/app_query.fasta_1.948
 -DB-GenEmbl -OFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosom62 -TRANS-human4.0.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=25 -MODE=LOCAL -OUTFMT=pfs -NORM-ext -MINLEN=0
 -MAXLEN=200000000 -USER=US09677752_@CGN1_1_6443 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-677-752-4
 Query length: 878
 Database: GenEmbl.*
 Database sequences: 1283235
 Database length: 1216004940
 Search time (sec): 2529.030000

score_list:

Sequence	Strd	Orig	zScore	Escore	Len	Documentation
gb_bal:AE001361	-	4509.00	5030.18	5.7e-272	5601	! AE001361 Chlamydia trachomatis
gb_bal:AE002294	+	3343.50	3941.98	2.3e-211	10908	! AE002294 Chlamydia muridarum
gb_bal:AE002192	+	947.50	1035.13	1.9e-49	10757	! AE001628 Chlamydia pneumoniae
gb_bal:AE002192	+	947.50	1033.40	2.4e-49	12676	! AE002192 Chlamydia pneumoniae
gb_bal:CPN133034	+	947.50	1030.13	3.6e-49	12780	! AJ133034 Chlamydia pneumoniae
gb_bal:AE002546	+	947.50	1000.01	1.7e-47	300650	! AP002546 Chlamydia pneumoniae
gb_bal:AE002546	+	940.50	1041.54	8.4e-50	2787	! AE002546 Chlamydia pneumoniae
gb_bal:AE002193	+	919.50	1004.52	9.7e-48	10026	! AE002193 Chlamydia pneumoniae
gb_bal:AE001627	+	915.50	995.75	3.0e-47	15068	! AE001627 Chlamydia pneumoniae
gb_bal:AE001841	+	906.50	1003.44	1.1e-47	2793	! AE001841 Sequence 15 from Patent
gb_bal:CPU72499	+	896.50	986.26	1.0e-46	4926	! U72499 Chlamydia pneumoniae
gb_bal:AE001829	+	883.00	977.05	3.3e-46	2815	! AE001829 Sequence 3 from Patent
gb_bal:AE002192	+	883.00	961.18	2.5e-45	12676	! AE002192 Chlamydia pneumoniae
gb_bal:CPN133034	+	883.00	957.91	3.8e-45	300650	! AJ133034 Chlamydia pneumoniae
gb_bal:AE002546	+	883.00	927.78	1.8e-43	10757	! AE002546 Chlamydia pneumoniae
gb_bal:AE001628	+	866.00	943.87	2.3e-44	10757	! AE001628 Chlamydia pneumoniae
gb_bal:AE001835	+	857.00	948.04	1.4e-44	2787	! AE001835 Sequence 9 from Patent
gb_bal:AE001851	+	853.50	943.34	2.5e-44	3000	! AE001851 Sequence 25 from Patent
gb_bal:CPN133035	+	853.50	935.98	6.4e-44	6030	! AJ001311 Chlamydia pneumoniae
gb_bal:AE001587	+	852.00	923.71	3.1e-43	16448	! AE001587 Chlamydia pneumoniae
gb_bal:CPN133035	+	852.00	918.52	6.0e-43	29520	! AJ133035 Chlamydia pneumoniae
gb_bal:AE001837	+	849.50	938.18	4.8e-44	3200	! AE001837 Sequence 1 from Patent
gb_bal:CPN133035	+	849.50	931.50	1.1e-43	6030	! AJ001311 Chlamydia pneumoniae
gb_bal:AE002235	+	845.50	919.65	5.2e-43	12127	! AE002235 Chlamydia pneumoniae
gb_bal:AE002545	+	845.50	885.82	4.0e-41	299650	! AP002545 Chlamydia pneumoniae
gb_bal:AE001837	+	844.50	934.15	8.1e-44	2757	! AE001837 Sequence 11 from Patent
gb_bal:AE001849	+	845.00	912.01	1.4e-42	2838	! AE001849 Sequence 23 from Patent
gb_bal:AE001629	+	817.00	889.06	2.6e-41	10695	! AE001629 Chlamydia pneumoniae
gb_bal:AE002191	+	817.00	884.95	4.4e-41	15789	! AE002191 Chlamydia pneumoniae
gb_bal:CPN133035	+	792.50	867.32	1.3e-40	6234	! U65943 Chlamydia pneumoniae
gb_bal:CPN133035	+	784.50	858.57	4.3e-39	6110	! U65943 Chlamydia pneumoniae
gb_bal:AE002293	+	784.00	826.51	8.0e-38	14482	! AE002293 Chlamydia muridarum
gb_bal:AE001360	+	758.00	820.02	1.8e-37	14177	! AE001360 Chlamydia trachomatis
gb_bal:AE001585	+	740.00	802.77	1.7e-36	10766	! AE001585 Chlamydia pneumoniae
gb_bal:AE002237	+	739.00	802.38	1.8e-36	10044	! AE002237 Chlamydia pneumoniae
gb_bal:AE001831	+	735.00	810.46	6.2e-37	3052	! AE001831 Sequence 5 from Patent
gb_bal:AE001586	+	708.50	767.68	1.5e-34	10574	! AE001586 Chlamydia pneumoniae
gb_bal:AE001833	+	705.50	779.42	3.3e-35	2526	! AE001833 Sequence 7 from Patent
gb_bal:AE002189	+	575.50	619.09	2.9e-26	10236	! AE002189 Chlamydia pneumoniae
gb_bal:AE001631	+	574.50	616.22	4.1e-26	12083	! AE001631 Chlamydia pneumoniae
em_bal:AB033816	+	566.00	621.96	2.0e-26	2844	! AB033816 Chlamydia pneumoniae

gb_bal:AE001638 + 566.00 604.58 1.8e-25 14781 ! AE001638 Chlamydia pneumoniae
 gb_bal:AE002181 - 566.00 601.12 2.9e-25 20517 ! AE002181 Chlamydia pneumoniae
 gb_bal:AP002547 - 566.00 572.81 1.1e-23 300550 ! AP002547 Chlamydia pneumoniae
 gb_bal:AE002190 - 546.00 585.55 2.1e-24 10746 ! AE002190 Chlamydia pneumoniae

seq_name: gb_bal:AE001361

seq_documentation_block:

LOCUS AE001361 5601 bp DNA BCT 30-OCT-2000
 DEFINITION Chlamydia trachomatis section 1 of 87 of the complete genome.

ACCESSION AE001361 AE001273

VERSION AE001361.1 GI:3329348

KEYWORDS Chlamydia trachomatis.

SOURCE Chlamydia trachomatis.

ORGANISM Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.

REFERENCE 1 (bases 1 to 5601)

AUTHORS Stephens,R.S., Kalman,S., Lammel,C.J., Fan,J., Marathe,R., Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q., Koonin,E.V. and Davis,R.W.

TITLE Genome sequence of an obligate intracellular pathogen of humans:

JOURNAL Science 282 (5389), 754-759 (1998)

MEDLINE 99000809

PUBMED 9784136

REFERENCE 2 (bases 1 to 5601)

AUTHORS Kalman,S., Mitchell,W., Lammel,C.J., Fan,J., Marathe,R., Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.

TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis

JOURNAL Nat. Genet. 21 (4), 385-389 (1999)

MEDLINE 99206606

PUBMED 10192388

REFERENCE 3 (bases 1 to 5601)

AUTHORS Stephens,R.S., Kalman,S., Lammel,C.J., Fan,J., Marathe,R., Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q., Koonin,E.V. and Davis,R.W.

TITLE Direct Submission

JOURNAL Submitted (20-MAY-1998) Program in Infectious Diseases, University of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA

FEATURES Location/Qualifiers

1..5601

/organism="Chlamydia trachomatis"

/strain="D/UW-3/CX"

/db_xref="taxon:813"

323..640

/gene="CT873"

323..640

/gene="CT873"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAC68471.1"

/db_xref="GI:3329349"

/translation="MSSNRMLATIKKTVLLSKRKAGIRIDYCALADAVEYSVLWF

LWGVYAVYSSGVREAIIEGETLSPGDAQDASLESAQVRLAGLMDYWDGDSKLECKK1"

complement(815..3451)

/gene="pmp1"

/note="CT874"

complement(815..3451)

/gene="pmp1"

/codon_start=1

/transl_table=11

/product="Putative Outer Membrane Protein I"

/protein_id="AAC68472.1"

/db_xref="GI:3329350"

/translation="MRPHMNFCCIAILSTAVILFQDPLIGETALTITKNNHVCT

FFEDCTMESLPALCAHASQDDPLVIGNSYCWVSKLHITDPKALFKEKGLDSION

PRFLSEFTSSKSSPSIITHOKGSLRNNGSMSCFARNHAGSGAISCADAFSLQHN

YLFTFEENSKNGGATQAOFTESLRVSPISFARNRADLNGAICCSNLTCISGNVN

PLFTFGNSATNGAICCSIDLNTSEKSLSLACNETLTFASNSAKEKGAIYAKHVL

RYNGPVSTINNSAKIGLTAIGESGSLIAGEGSLVLFQNNQSRTSDOGLVRNAIYLE

KDAILLSLENRGDIILFDPIVQESSKESPLPSLQASVTSPTPATASPLVIQTSAN

434 aLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeuLeuIleMet 450
|||||
2151 TCCTTTCCGGCGCTCTCTCTCAGAGCTCTCAGCTCTCCTCATATG 2102
451 GluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAlaThrLe 467
|||||
2101 GAAGCGGAACTTCTTTAAAACTTCTCTCTGATTTCGAAGTTAGCTACGCT 2052
467 uSerIleProLeuHisSerLeuAspThrGluLysSerValThrIleHisA 484
|||||
2051 AAGTATTCCTCTCAATTCCTTAGATACGTAAGAAAGCGTAACATATCCAG 2002
484 laProAsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspGlu 500
|||||
2001 CCCCTAACCTTTCTATCCAAAGATCTCTCTCTCTAATCTGGAGATGAG 1952
501 AsnPheTyrGluAsnValGluLeuSerLysGluGlnAsnIlePr 517
|||||
1951 AATTTTATCAAAATGTAGAGCTTCTCAGTAAAGAGCAAAACAATATCC 1902
517 oLeuLeuThrLeuSerLysGluGlnSerHisLeuHisLeuProAspGlyA 534
|||||
1901 TCTCCTTACTCTCTCTAAAGAGCAATCTCATTTACATCTTCTCTGATGGA 1852
534 snLeuSerSerHisPheGlyTyrGlnGlyAspThrPheSerTrpLys 550
|||||
1851 ACCTCTCTTCTCACTTTGGATATCAAGAGATGGACTTTTCTTGGAAA 1802
551 AspSerAspGluGlyHisSerLeuIleAlaAsnTrpThrProLysAsnTy 567
|||||
1801 GATTCTGATCAAGGGCATCTCTGATTGCTAATTTGGACGCTAAAACTA 1752
567 rValProHisProGluArgGlnSerThrLeuValAlaAsnThrLeuTrpA 584
|||||
1751 TGTGCTTCATCCAGAACGCTCAATCTACACTCGTTTGGAACTCTTTGGA 1702
584 snThrTyrSerAspMetGlnAlaValGlnSerMetIleAsnThrIleAla 600
|||||
1701 ACACCTATTCGGATATCAAGCTGTGAGTGTGATGATTAATACANTAGCG 1652
601 HisGlyGlyAlaTyrLeuPheGlyThrTrpGlySerAlaValSerAsnLe 617
|||||
1651 CACGAGGAGCCATCTATTTTGGAACTGGGATCTGCTGTTCTAATTT 1602
617 uPheTyrAlaHisAspSerSerGlyLysProIleAspAsnTrpHisHisA 634
|||||
1601 ATCTATGCTCACGACAGCTCTGGAAACCTATTCGATAATTTGGCATCAT 1552
634 rgSerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAspHis 650
|||||
1551 GAAGCCCTGGCTACCTATTCGGTATCAGTACTACAGTTTAGATGACCAT 1502
651 SerPheCysLeuAlaGlyGlnLeuLeuGlyLysSerSerAspSerPh 667
|||||
1501 TCTTCTGCTTGGCTGACGACAAATTAATCTCGGAAATCTGCTCGATTCCT 1452
667 eileThrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaGlnL 684
|||||
1451 TATTACGCTACAGAAACGACCTCTATATAGCTACTCTACACGCGCAAC 1402
684 euAlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrAsnGluSer 700
|||||
1401 TCGTACCCCTCTTAATGAAATCTCTGCACAGGATGCTAATATGAAGT 1352
701 ileHisGluLeuLysThrLysTyrArgSerPheSerLysGluGlyPheG 717
|||||
1351 ATCCATGAGCTAAACAAATATCGCTCTCTCTAAAGAGGATTCGG 1302
717 ySerTrpHisSerValAlaValSerGlyGluValCysAlaSerIleProI 734
|||||
1301 ATCTGTCATAGCTGCTCAGTATCCGAGAGAGTGTGCGCATCGATTCCTA 1252

734 leValSerAsnGlySerGlyLeuPheSerSerPheSerIlePheSerLys 750
|||||
1251 TTGTATCCATGTTCCGGACTGTTCCAGCTCTTCTCTATTTTCTCTAAA 1202
751 LeuGlnGlyPheSerGlyThrGlnAspGlyPheGluGluSerSerGlyG 767
|||||
1201 CTCGAAGGATTTTCAGAACACAGGACGGTTTTCAGGAGAGTTCCGGAGA 1152
767 uileArgSerPheSerAlaSerSerPheArgAsnIleSerLeuProIleG 784
|||||
1151 GATTGGTCTCTTTCGCCAGCTCTTTCAGAAATATTTCACTTCCCTATGG 1102
784 lyIleThrPheGluLysLysSerGlnLysThrArgThrTyrTyrPhe 800
|||||
1101 GAATAACATTTGAAAAAATCCAAAAACACGAAACTACTATTACTTT 1052
801 LeuGlyAlaTyrIleGlnAspLeuLysArgAspValGluSerGlyProVa 817
|||||
1051 CTGGAGCCTTACATCCCAAGACCTAAAAACGTGTGTGGAATCGGACCTGT 1002
817 lValLeuLeuLysAsnAlaValSerTrpAspAlaProMetAlaAsnLeuA 834
|||||
1001 AGTGTACTCAAAATGCCGTCTCTGGGATGCTCCTATGGCGACTGG 952
834 spSerArgAlaTyrMetPheArgLeuThrAsnGlnArgAlaLeuHisArg 850
|||||
951 ATTCGCGAGCTTACATGTTCCAGGCTTACGAATCAAGAGCTCTGCATAGA 902
851 LeuGlnThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSe 867
|||||
901 CTTTCAGACGCTGTTAAATGTCTCTACCTACTGCGCGGCAAGCCATAG 852
867 rTyrSerLeuAspLeuGlyThrThrTyrArgPhe 878
|||||
851 TTACTCCTGGATCTGGGACCACTTTACAGGTTT 818

seq_name: gb_ba1:AE002294

seq_documentation_block: 10908 bp DNA BCT 26-MAY-2000
LOCUS AE002294
DEFINITION Chlamydia muridarum, section 26 of 85 of the complete genome.
ACCESSION AE002294 AE002160
VERSION AE002294.1 GI:7190305
KEYWORDS
SOURCE Chlamydia muridarum.
ORGANISM Chlamydia muridarum
REFERENCE 1 (bases 1 to 10908)
AUTHORS Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Guinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
TITLE Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
JOURNAL Nucleic Acids Res. 28 (6), 1397-1406 (2000)
MEDLINE 20150255
PUBMED 10684935
REFERENCE 2 (bases 1 to 10908)
AUTHORS Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Guinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1..10908
/organism="Chlamydia muridarum"
/db_xref="taxon:83360"
/note="synonym: Chlamydia trachomatis MoPn"
241..495
gene

CDS	complement(9731..10834) /gene="rc0273" /note="conserved hypothetical protein; identified by Glimmer2: putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAF39141.1" /db_xref="GI:7190313" /translation="MVKIMAPITPTSPQVKGLLSRFLTADPHPKLRYVYDISLIAI STLCIVSIILWTOGSGLAFALPAALCALGVTLIVSDIAESPKSEVADPTAAVSL PFILGTAGLAFMSAIVGGGAVILANPLFLMGSLFALMSLHKVTYQLNSRSOW OKONIKIOEASWENKLPKESKESLQTSYSLARKDKTRRNKPMGNKGSQVPA SIANTERLSRFEVLHSSLLRQKELFNTSNIKKELPNTKSLILHTPLNRRSPSGSDS DDVYTPRAGSSATGALGDISGSSSTSTSTPKAKRRVVRSSERNARHRRN KEDHPRNOEESDDSDSLPLSPRRKKYRSRPK"
BASE COUNT	3210 a 1860 c 2587 g 3251 t
ORIGIN	
alignment_scores:	Quality: 3543.50 Length: 879 Ratio: 4.429 Gaps: 2 Percent Similarity: 91.013 Percent Identity: 77.816
alignment_block:	US-09-677-752-4 x AE002294/rev ..
Align seg 1/1 to reverse of: AE002294 from: 1 to: 10908	
1 MetArgProAspHisMetAsnPheCysCysLeuCysAlaAlaIleLeuSe 17	183 heSerLeuSerArgAsnValSerProIleSerPheAlaArgAsnArgAla 199
3707 ATGGACAGCATCATGTGAACCTTATGTTGCTATGCGCACATTCCTC 3658	3157 TTTCCTTTGCAAGAAATGCTCTCTCTCTTTTCCCGTAATCGCGCA 3108
17 rSerThrAlaValLeuPheGlyGlnAspProLeuGlyGluThrAlaLeuL 34	200 AspLeuAsnGlyGlyAlaIleCysCysSerAsnLeuLeuCysSerGlyAs 216
3657 CCCACACGACATCTCTTTGTCAGGATGCGTTAGATAAATCCGCTCTTA 3608	3107 AATCTGAATGGAGGAGCTATTTCGCCAAAATCTCATTTGTTTCAGGAAA 3058
34 euThrLysAsnProAsnHisValValCysThrPhePheGluAspCysThr 50	216 nValAsnProLeuPhePheThrGlyAsnSerAlaThrAsnGlyGlyAlaI 233
3607 TTACTAAAAATCCTAACAGCATCGTCTGTACATTTTGGAAAGACTGCACC 3558	3057 TGTAAACCCCTCTTTTTCACCAATACTCTGCCCTAAATGGAGGAGCTA 3008
51 MetGluSerLeuPheProAlaLeuCysAlaHisAlaSerGlnAspAspPr 67	233 leCysCysIleLeuAsnThrSerGlyLysGlySerLeuSerLeu 249
3557 ATGGGAACCTTTCTCCCTGCTCTCTCTCATGCACGACAGATGATCC 3508	3007 TTTGTTGTATTAATGAACAAATCTCTCAGAAAAGGATCGCTCTCCTTG 2958
67 OleuTyrValLeuGlyAsnSerTyrCysTrpPheValSerLysLeuHisI 84	250 AlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysGluLysG 266
3507 GCTATACATAATAGGAATACCCATAACTGGTTCGTATCCAAATCCACCC 3458	2957 GCATATAACCAAGAGACTCTATTTCAGGCAATTCAGCGAAGGAAAAGG 2908
84 leThrAspProLysGluAlaLeuPheLysGlyLysGlyAspLeuSerile 100	266 yGlyAlaIleTyrAlaLysHisMetValLeuArgTyrAsnGlyProValS 283
3457 CCTCAACCAATCAAGACGGTTTCTTAAAGAAAAGGCGATCTTTCTATT 3408	2907 GGGAGCCATTACACAAACATATGTTTTCGGTCACAAATGGACCCGTTT 2858
101 GlnAsnPheArgPheLeuSerPheThrAspCysSerSerLys...GluSe 116	283 erPheIleAsnAsnSerAlaLysIleGlyAlaIleAlaIleGlnSer 299
3407 CAAGACTTTTCGATTTCTTTCATTTACGGATTCCTCTCTCATGCACGAC 3358	2857 CTTTCGTAATAACAGTCTAAATAGGGGAGCTATTGCTATCCAGCTCA 2808
116 rSerProSerIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnA 133	300 GlyGlySerLeuSerIleLeuAlaGlyGluGlySerValLeuPheGlnAs 316
3357 CTCCTCTCTATCTCTACCATAAAATGGGAGCTATTCCTAAGAATA 3308	2807 GGAGGAAGTCTCTCCATCATTCAGGGGAGGCTCTGTCTCTTTTCCAAAA 2758
133 snGlySerMetSerPheCysArgAsnHisAlaGluGlySerGlyGlyAla 149	316 nasSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrL 333
3307 ATGGAAACATGAGCTTTTATCGCAATCATTCGCAAGGCTCTCGAGGAGCA 3258	2757 TAATCTCTGTCATTCTCTGATCAAGGTACGGTAAGAAATCGGATTACT 2708
150 IleSerAlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaph 166	333 euGluLysAspAlaIleLeuSerSerLeuGluAlaArgAsnGlyAspIle 349
3257 CTCCTCCACAGCGCATTTGTTCTTACAAACAACAATATCTTTTTCACAAATTT 3208	2707 TAGAGAAAATGCTCTGCTCTCTCTTTTGGAAAGTCCGCCATGGGATATC 2658
166 eGluGluAsnSerSerLysGlyAsnGlyGlyAlaIleGlnAlaGlnThrP 183	350 LeuPhePheAspProIleValGlnGluSerSerLysGluSerProLe 366
3207 TGAAGAAAATTTCTTCTGCAAAAATGGCGGTGCCATTCAGCTCAACATC 3158	2657 CTTTCTTTTGATCCATTGTGCAAGAGTCTGTAAGTCTCTGAATTTCTTAC 2608

IQOLVATKQVROSETRIGWCEGINSFPHKDKTKINKGRHISAGYVVGATTTLASDNL
 ITAFPCOLGKDRDHFINKNRASAYASLHLHLATLSSPSLLRYLPGSSESEPVLPD
 AQISYIYSKNTMTYITOAQPGSSWYNDCCALELASSLPHALSHGLPHAFPPFK
 VEASYTHODSFKERNITLVRSFDSGLINVSCELTIFPERSRNRSYATATVIYAD
 VYRKNPDCTTALLINNTSWKTTGTLNLSKQAGIRAGIFYAFSPNLEVTSLNLSMEIRGS
 SRSYNADLGKGFQF
 complement(3233. .4348)
 /gene="yxjg_2"
 /note="Cp0448"
 complement(3233. .4348)
 /gene="yxjg_2"
 /codon_start=1
 /transl_table=11
 /product="Hypothetical Protein"
 /protein_id="AAD18592.1"
 /db_xref="GI:4376732"
 /translation="MEMSPFPQPEOCHDFVVGFLRPESLTRARSDFEGRIVYEQM
 RVVEDAAIRNLIKKQTEAGLIFFTDGEFRYSDFMWGFHGVDRDRSDNPDIQVY
 LKDISYKPKFIEHFEVKTIEKGNKAKOTIPSPQFHEMIFAPNLKNTKRFYPT
 NQELIDIVFYVROVIODLVAAGCRNLQDDCWCRLDIRAPSWGVDSHRLQEL
 EQFLWHLNVMKDRPDLFVSLHVCRDGYQAEFFSRAYDSIEPLAKTDVDSYHY
 WALDDKYSGGAEPYVSGEKHVCGLIISNHSICIEDORDVVSRIYEAASYIPIERUS
 LSPQCGFASCGDHRMTTEEQWKKIAFVKEIAKEIWG"
 complement(4865. .6769)
 /gene="pmp_10.1"
 /note="Cp0449"
 complement(4865. .7650)
 /note="Polymorphic Outer Membrane Protein G Family;
 contains frame shift"
 complement(6715. .7650)
 /gene="pmp_10.2"
 /note="Cp0450"
 7813. .10599
 /gene="pmp_11"
 /note="Cp0451"
 7813. .10599
 /gene="pmp_11"
 /codon_start=1
 /transl_table=11
 /product="Polymorphic Outer Membrane Protein G Family"
 /protein_id="AAD18593.1"
 /db_xref="GI:4376733"
 /translation="MKTSPVLYSVSLFASCHLOLANEELLSPDDFNGNIDSGTF
 TPKTSATYSLTGDFEFPCKGTPLSDFCKQTFDNLTLFLNGHSLTFEGFIDACTHA
 GAAASTTANKNLTFSGFSLSDSPSTVITGQTLSSAGVNLNIRKLVVAGNFS
 TADGAIKAGFLLTGTSGDALFSNNSTKGAATATTAGARIANNITGVRFSLNIA
 TSGGAIDDEGTSILSNKFLYFEGNAKTGCAICNTRKASGPELIIINNKTLIFASN
 VAETSGGAIHAKKALSSGGTFEFURNVSSATPKGGAISIDASGELSLSAETNITF
 VNTLTGSDTDPKRNAINIGSKGFTELRAKNHTIFFYDPTITSEGTSSDLVKINN
 GSAGALNPQGTILFSGETLADDELKVADNLKSSFTQPVSLGGKLLLOKGVLETS
 FSQEAGSLGMDSGTLLSTAGSITITNLGINVDLQKQVSLTAKGASKNVIVSGK
 LNLIDIEGNISHEMSHDLFSLKITVDADVNDVIDISLIPVPAEDPNSEYGFQ
 QNWNMTDTATNTKEATATWTGTGVPSPERKSALVCNTLMGVTDIRLQQLVEIG
 ATGMCHKOGFWSMTNLTHTKGTGDNKRGFRHTSGYVIGGSAHTPKDGLTFACFL
 FARDKDCFIHNNSTYGTTLFKHSHTLOPNYLRIGRAKFSESAIEKFEIPEILAL
 DVQSFSDNRMETHYSLPSEGSWSNEIAGIGLDLPVFLNPHIPLFTFIPQM
 KYVMYVNSONSFDSSDGRGFSIGRLNLSIPVGAKEVQGDIGDSYTDLSGFPVSO
 VYRNPNQSTALIVMPSDKIRGNLSRQAFLLRGSNNVYVNSNCELFCYHAMELRGS
 SRYNNVDVGTKLRF"

BASE COUNT 3228 a 2459 c 2065 g 3005 t
 ORIGIN
 alignment_scores:
 Quality: 947.50 Length: 910
 Ratio: 1.771 Gaps: 32
 Percent Similarity: 58.791 Percent Identity: 31.648
 alignment_block:
 US-09-677-752-4 x AE001628 ..
 Align seg 1/1 to: AE001628 from: 1 to: 10757

85 ThrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerIleol 101
 ||||| : : : : : ||||| : : : : :
 259 ACAATGCAGATGGAACATATCTATAATCTAACAGGGATGTCTCAATCAC 308
 ||||| : : : : : ||||| : : : : :
 101 nAsnPheArgPhe...LeuSerPheThrAspCysSerSerLysGluSerS 117
 ||| : : : : : ||||| : : : : :
 309 CAATGCAGGATCTCCGACAGCTCTAACCGCTTCCTCTTAAAGAAACTA 358
 ||||| : : : : : ||||| : : : : :
 117 erProSerIleIleHisGlnLysAsnGly...GlnLeuSerLeuArgAsn 132
 ||||| : : : : : ||||| : : : : :
 359 CTGGGAATCTTTCTTCCAGGCCAGGCTACCAATTTCTCTACAAAT 408
 ||||| : : : : : ||||| : : : : :
 133AsnGlySe 135
 ||||| : : : : :
 409 ATCGATCGGGAGGAGCACTCTACCTTTACCAATACAGCTGCCAATAAGCT 458
 ||||| : : : : :
 135 rMetSerPhe.....CysArgAsnH 142
 ||||| : : : : :
 459 TCTCTCTTTTTCAGGATCTCTCTATTGTCTAATAACAAACCAAGATG 508
 ||||| : : : : :
 142 isAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSerLeuGln 158
 ||||| : : : : : ||||| : : : : :
 509 CTACCACAGAACAGGAGCATCAAGTCCACAGGAGCTTGTCTATTCAG 558
 ||||| : : : : : ||||| : : : : :
 159 HisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerLysGlyAsnG 175
 ||||| : : : : : ||||| : : : : :
 559 TCGAATAT...AGTTGCTACTTTGGCCAAAACCTTTTCTAATGACAAATG 605
 ||||| : : : : :
 175 yGlyAlaIleGlnAlaGlnThrPheSerLeuSerArgAsnValSerProI 192
 ||||| : : : : : ||||| : : : : :
 606 AGCGCGCTCCAGGAGGCTCTATCATCTATCGCTATAAACCCCAAC...C 652
 ||||| : : : : : ||||| : : : : :
 192 leSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCysCys 208
 ||||| : : : : : ||||| : : : : :
 653 TAACGTTTGGCAAAAACAAAGCAACGCAAGGGGGTGCCTCTATTCC 702
 ||||| : : : : : ||||| : : : : :
 209 serAsnLeuIleCysSerGlyAsn...ValAsnProLeuPhePheThrG 224
 ||||| : : : : : ||||| : : : : :
 703 ACGGAGGGGATTACAATTAAACAATACGTTAAACTCAGCATCATTTTCTGA 752
 ||||| : : : : : ||||| : : : : :
 224 yAsnSerAlaThrAsn..... 229
 ||||| : : : : : ||||| : : : : :
 753 AAATACCGCGCGCAACAATAGCGGAGGCAATTTACACGGAAGCTAGCAGTT 802
 ||||| : : : : : ||||| : : : : :
 229 229
 803 TTATTACGACACACAAGCAATAGCTTTTATAACAATAGTGTGACCGCA 852
 ||||| : : : : : ||||| : : : : :
 230GlyGlyAlaIleCysCysIleSerAspLeuAsnThrSe 242
 ||||| : : : : : ||||| : : : : :
 853 ACCTCAGCTACAGGGGAGGCAATTTACTGTAGTAGT.....ACATCAG 896
 ||||| : : : : : ||||| : : : : :
 242 rGlyGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAlaG 259
 ||||| : : : : : ||||| : : : : :
 897 CCCCACCAACAGCTCTTAACCTATCATCAGACACGCGGGAACCTGAAC 946
 ||||| : : : : : ||||| : : : : :
 259 erAsnSerAlaLysGluLysGlyAlaIleTyrAlaLysHisMetVal 275
 ||||| : : : : : ||||| : : : : :
 947 GAAATACAGCAATTTACTAGTGGTGGGCGATTATATCTGACCAATCTAGT 996
 ||||| : : : : : ||||| : : : : :
 276 LeuArgTyrAsnGlyProValSerPheIleAsnAsnSer..... 288
 ||||| : : : : : ||||| : : : : :
 997 CTTTCCTCTCGAGGACCTACCGCTTTTAAAAACAACACTCTGCTATAGATAC 1046
 ||||| : : : : : ||||| : : : : :
 289AlaLysIleGlyGlyAlaIleAlaIleGlnSerGlyGlySerLeuS 304
 ||||| : : : : : ||||| : : : : :
 1047 TGCAGCTCCCTTAGGAGGAGCAATTCGCAATTCGTGACTCTGGATCTTTGA 1096
 ||||| : : : : : ||||| : : : : :
 304 erIleLeuAlaGlyGluGlySerValLeuPheGlnAsnAsnSer..... 318
 ||||| : : : : : ||||| : : : : :
 1097 GTCTTTCGGCTCTGGTGGGAGACATCACTTTTGAAGGAAACACACATGTC 1146
 ||||| : : : : : ||||| : : : : :
 319 ...GlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeu... 333

seq_name: gb_bal:AE002192	gene
seq_documentation_block:	CDS
LOCUS AE002192 12676 bp DNA BCT 30-MAY-2000	
DEFINITION Chlamydomydia pneumoniae AR39, section 25 of 94 of the complete genome.	
ACCESSION AE002192 AE002161	
VERSION AE002192.2 GI:8163405	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
PUBMED	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
FEATURES	
source	
1..12676	
/organism="Chlamydomydia pneumoniae AR39"	
/strain="AR39"	
/db_xref="taxon:115711"	
/note="synonym: Chlamydia pneumoniae AR39"	
complement(90..185)	
/gene="CP0300"	
complement(90..185)	
/gene="CP0300"	
/note="hypothetical protein; identified by Glimmer2; putative"	
/codon_start=1	
/transl_table=11	
/product="hypothetical protein"	
/protein_id="AAF38157.1"	
/db_xref="GI:7189227"	
/translation="MFTSLGRSAIFFTCSLKHRLRIGFELEIFQR"	
complement(275..1819)	
/gene="CP0301"	
complement(275..1819)	
/gene="CP0301"	
/note="similar to GP:4376735; identified by sequence similarity; putative"	
/codon_start=1	
/transl_table=11	
/product="polymorphic membrane protein G family"	
/protein_id="AAF38158.1"	
/db_xref="GI:7189228"	
/translation="MTILRNFLTCSALFLALPAAQVYVYVHSDGYNGAINNKSLEPK	
ITCYEGTSYIFLDVIRISNVKHDOEACVFNRSNGLFPMGNRCNFTFHNLMTEGFG	
AAISRYGDTTLTLNFSYLAFTSAPLLPQGGGAYSLGSVMIESEETFCGNYSSW	
SGAAYTPYLLKASRPVNLUSGNRLVFRDNVSGYIGGAIKSTNLTITTRGSCFE	
NNHAYDVNSNGAIAIPGSSISVKSVDLIFKGTASQDGNTHNSILHQSQAQF	
KNLRAVSESGYFYDPIHSHSESHKITDLVINAPKEGTVEGTISFSGCLDDHEVCAE	
NLTSTILODTLWAGTSLSDGVTIQLHFKFOEASSTLWSPGTTLLCSGDARVNLH	
ILIEDTNPVPIRAEDKDALSVLEKLVAFAYKWSYVDFPQKEAFTIPLELLGP	
SFDSLLLETTTLRTQVTTENDAVRGFWSLSWEEYPPSLDKDRITPKTKTVFLTNWP	
EITSTP"	
complement(2077..4926)	
/gene="CP0302"	
complement(2077..4926)	
/gene="CP0302"	
/note="similar to GP:4376733; identified by sequence similarity; putative"	
/codon_start=1	
/transl_table=11	
/product="polymorphic membrane protein G family"	
/protein_id="AAF38159.1"	
/db_xref="GI:7189229"	
/translation="MIYLCFCYIDANSLKNKSNITMTSIPWLVSSVLAFLSCHLQSL	
ANEALLPDDSFNGNIDSTPTKTSATYSLTGDFVEFEPKGTPLSDSCFKOTTDN	
LTFLNGHSLTFEIDAGTHAGAASTANKNLTFSGLSFSSPSTVTYTGQGLT	
SSAGVNLNIRKLVAGNFSTADGGAIKGASFIILTGTSGDALFSNNSSTTKGAAT	
TAGARIANTGYVRFSLNASTSGAIDDECTSLNNKFLYFECNAKTTTGGACINT	
KASGPGLIISNNKTLPASNVAETSGGAIHAKKLLALSSGGTBEFLRNVSATPKGG	
AISDASGLSLSAETGNITFVRNTLTITSTDTPKRNAINIGSNGKFEFLRAAKNHT	
IFFYDPTISGTSDDVLKINNGSAGALNPYQGTILFSGETLTADKLKADNLSKSTQ	
PVSLGGKLLQKGVLESTFSQEAAGLLGMDSGTTLSTAGSTITNLTGADVDTNV	
LKQVSLTAKGASNKVIVSGKLNLDIEGNIYESHMFSDQLFSLKLTVPDADTAV	
DISLIPVPAEDPNSEYGFQGVNNTDTATNKEATATKTGTFVPSRERKSLV	
CNTLWGVFTDRIQQQLVEIGATGMEHKQGFVSSMTNFKHTGDNKKRGRHRTSGGY	
VIGSAITPKDGLTFEAFCHLFARDKDCEIAHNNSTYGGTLFFKHSHLTQPNYLR	
GRKFSSEAIEKFEIPREIPLADVOVFSHSDNRMETHYTSLSPEGSWSNCSIAAGG	
LDPFLVLSNPFLPKTEIPOMKVMVYVSONSFESSDGRGFSIGRLNLSIPVGA	
FVQDNCDSYTYDLSSGPFVSDVYRNPOSTATLVMSPDMSKIRGNLSRQAFLLRGSN	
NYVNSNCELFGYAMELURSSRNYNVDVGTKLRF"	
5026..7812	
/gene="CP0303"	
5026..7812	
/gene="CP0303"	
/note="similar to GP:4376729; identified by sequence similarity; putative"	
/codon_start=1	
/transl_table=11	
/product="polymorphic membrane protein G family"	
/protein_id="AAF38160.1"	
/db_xref="GI:7189230"	
/translation="MKSQFSLVLSTLACFTSCSTVFAAETNIPGSDSPGDSFTNTG	
TYTPKNTTGTIDYTLTGDTLQNLGDSAAITKGCDFSTESLSFAGKGYSLFLNKS	
SAEGAAALSVTTDKNLSTGFSFLAAPSIVITPSSGKAVKCGDLTFDNNGTILF	
KODYCEENGGAISTKNLSKNGTSISFEGNSKSSATGKKGAIKATGTVDTNNTAPT	
LFSNIAEAGGAINSTGNCITIGTSLVFSNSVATAGNAGIAGSDADVTISGNOS	
VTFSGNAVANGGAIYAKKLTILASGGGGISFNSNIVOGTTAGNAGSAILAAGCSL	
SAEAGDITFNGNAIVATTPOTTKRNSIDIGSTAKITNLRAISCHSIFTFDPTANTAA	
DSTDITNLNKADAGNSTDYSGSIVFSGKSEDEAKVADNLTSLIKQPVTLTAGNLVL	
KRGVLDTKFTGTAGSSVIMDAGTTLKASTEEVLTGLSIPVSLGKGVYIAASA	
ASKNVALSGPILLDNQNAVYENHDLGTQDFSVQLSALGTATTDVPAVPTVATPT	
HYGYGTWGTWVDDTSTPKTKTATLAWNTGYLPNPEROGPLVNSLGSFSDIOA	
IQGVIERAALTLCSDRGVAAGVANFLDKDKKGEKXKVRHKSGGVAIGGAACSENL	
ISFACOLFGSKDKFLVAKNHTDYTAGFYIOHITCSGFIGCLLDKLPKGSWKPLV	
LEGOLAYSHVSNDLAKTYTAYPEVKSGMNNFNMMLGASHSHSYPEYLHCFDTYAPYI	
KLNLTYIRODSESEKGTGRSFDSDNLFNLSLPIGVKFEKFDNDNFSDYDLTSLVYPD	
LIRNDPKCTTALVISGASWETVANNLARQALQVRAGSHVAFSPMFEVLQGFVEVRGS	
SLIYNVDLGGKRFQ"	
complement(8170..8292)	
/gene="CP0304"	
complement(8170..8292)	
/gene="CP0304"	
/note="hypothetical protein; identified by Glimmer2; putative"	
/codon_start=1	
/transl_table=11	
/product="hypothetical protein"	
/protein_id="AAF38161.1"	
/db_xref="GI:7189231"	
/translation="MDLSRWLNKLKTRKQLQDNQNKYDKCKSYCLRGISQILGWV"	
8329..9444	
/gene="CP0305"	
8329..9444	
/gene="CP0305"	
/note="conserved hypothetical protein; identified by	

us-09-677-752-4.rge

```

679 hrValGlnAlaGln...LeuAlaThr.....SerLeuMetLys 690
      : : : : : ||| ||||| |||||
10443 CTCCTCATCTCCAGCATCTAGCGACCTTGTCTCTCCAAGCTTCTTACCG 10394
      : : : : : ||| ||||| |||||

691 .....lleSerAlaGlnAl 695
      : : : : : ||| ||||| |||||

10393 TACCTTCTCGGATCTGAAGGTGAGCAGCGCTGCTCTTTGATGCTCAGAT 10344

695 aCysTyrAsnGluSerIleHisGluLeuLysThrLysTyrArgSerPheS 712
      ||| ||| : : : : : ||||| ||| : : :
10343 CAGCTATATCTATAGTAAATACTATCAAAACCTATTATCACCAAGCAC 10294

712 erLysGluGlyPheGlySerTrpHisSerValAlaValSerGlyGluVal 728
      ||| ||| : : : : : ||||| ||| : : :
10293 CAAG...CGAGAGAGCTTGGTGTATAATGACGGTTGGCTCTCGAACTT 10247

729 CysAlaSerIleProIleValSer...AsnGlySerGlyLeuPheSerSe 744
      : : : : : ||||| : : : : : ||||| ||||| : : :
10246 GCGAGCTCCCTACCACACACTGCTTTAAGCCATGAGGCTCTCTTCCAGCG 10197

744 rPheSerIlePheSerLysLeuGlnGlyPheSerGlyThrGlnAspGlyP 761
      : : : : : ||| ||||| : : : : : ||||| ||||| : : :
10196 GTATTTCCTTTCATCAAAAGTAGAAGCTTCGTACATACACCAAGATAGCT 10147

761 heGluGlu...SerSerGlyGluIleArgSerPheSerAlaSerSerPhe 776
      : : : : : ||| ||||| : : : : : ||||| ||||| : : :
10146 TCAAAAGACGTAAATACTACTACCTTGGTACGACTTTCATAGCGGTGATTTA 10097

777 ArgAsnIleSerLeuProIleGlyIleThrPheGluLysLysSerGlnly 793
      ||||| ||||| : : : : : ||||| ||||| : : : : : ||||| : : :
10096 ATTACGCTCTCTGCGCTATTGCAATTACCTTCGAGAGATTCTCGAGAAA 10047

793 sThrArgThrTyrTyrPheLeuGlyAlaTyrIleGlnAspLeuLysA 810
      : : : : : ||| ||| : : : : : ||||| : : : : : ||| : : :
10046 CGAGCGTGGCTTTACCAAGACTACTGTCATCTACGTTGCCGATGCTATCT 9997

810 rGAspValGluSerGlyProValValLeuLeuLysAsnAlaValSerTrp 826
      : : : : : ||| ||| : : : : : ||||| ||| : : : ||||| : : :
9996 GTAAGAATCCTCATGCTGCAGCAGCTCTCTCTATCAACAATACCTCGTGG 9947

827 AspAlaProMetAlaAsnLeuAspSerArgAla..... 837
      : : : : : ||| ||||| : : : : : ||||| : : : : : ||| : : :
9946 ANAACTCAGGAACGAATCTCTCAAGACAAGCTGGTATCGGAAGAGCAGG 9897

838 .....TyrMetPheArgLeuThrAsnGlnArgAlaLeuHisArgLeuG 852
      ||| ||| : : : : : ||||| : : : : : ||||| : : : : : ||| : : :
9896 GATCTTTTATGCTCTCTCT.....CCAAATCTTGTG 9868

852 lThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSerTyr 868
      : : : : : ||||| : : : : : ||||| ||||| : : : : : ||||| : : :
9867 AGGTCAACAGTACACCTATCTATCGAAATTCGTGGATCTTTCACGCCACTAC 9818

869 SerLeuAspLeuGlyThrThrArgPhe 878
      : : : : : ||||| ||||| : : : : : ||||| : : : : : ||||| : : :
9817 AATGCAGATCTTGGAGCTGAAGTTCCAGTTC 9788

```

seq name: ab ba2:CPN733034

seq documentation block:

seq_locations

511201
511202
511203
511204
511205
511206
511207
511208
511209
511210
511211
511212
511213
511214
511215
511216
511217
511218
511219
511220
511221
511222
511223
511224
511225
511226
511227
511228
511229
511230
511231
511232
511233
511234
511235
511236
511237
511238
511239
511240
511241
511242
511243
511244
511245
511246
511247
511248
511249
511250
511251
511252
511253
511254
511255
511256
511257
511258
511259
511260
511261
511262
511263
511264
511265
511266
511267
511268
511269
511270
511271
511272
511273
511274
511275
511276
511277
511278
511279
511280
511281
511282
511283
511284
511285
511286
511287
511288
511289
511290
511291
511292
511293
511294
511295
511296
511297
511298
511299
511300
511301
511302
511303
511304
511305
511306
511307
511308
511309
511310
511311
511312
511313
511314
511315
511316
511317
511318
511319
511320
511321
511322
511323
511324
511325
511326
511327
511328
511329
511330
511331
511332
511333
511334
511335
511336
511337
511338
511339
511340
511341
511342
511343
511344
511345
511346
511347
511348
511349
511350
511351
511352
511353
511354
511355
511356
511357
511358
511359
511360
511361
511362
511363
511364
511365
511366
511367
511368
511369
511370
511371
511372
511373
511374
511375
511376
511377
511378
511379
511380
511381
511382
511383
511384
511385
511386
511387
511388
511389
511390
511391
511392
511393
511394
511395
511396
511397
511398
511399
511400
511401
511402
511403
511404
511405
511406
511407
511408
511409
511410
511411
511412
511413
511414
511415
511416
511417
511418
511419
511420
511421
511422
511423
511424
511425
511426
511427
511428
511429
511430
511431
511432
511433
511434
511435
511436
511437
511438
511439
511440
511441
511442
511443
511444
511445
511446
511447
511448
511449
511450
511451
511452
511453
511454
511455
511456
511457
511458
511459
511460
511461
511462
511463
511464
511465
511466
511467
511468
511469
511470
511471
511472
511473
511474
511475
511476
511477
511478
511479
511480
511481
511482
511483
511484
511485
511486
511487
511488
511489
511490
511491
511492
511493
511494
511495
511496
511497
511498
511499
511500
511501
511502
511503
511504
511505
511506
511507
511508
511509
511510
511511
511512
511513
511514
511515
511516
511517
511518
511519
511520
511521
511522
511523
511524
511525
511526
511527
511528
511529
511530
511531
511532
511533
511534
511535
511536
511537
511538
511539
511540
511541
511542
511543
511544
511545
511546
511547
511548
511549
511550
511551
511552
511553
511554
511555
511556
511557
511558
511559
511560
511561
511562
511563
511564
511565
511566
511567
511568
511569
511570
511571
511572
511573
511574
511575
511576
511577
511578
511579
511580
511581
511582
511583
511584
511585
511586
511587
511588
511589
511590
511591
511592
511593
511594
511595
511596
511597
511598
511599
511600
511601
511602
511603
511604
511605
511606
511607
511608
511609
511610
511611
511612
511613
511614
511615
511616
511617
511618
511619
511620
511621
511622
511623
511624
511625
511626
511627
511628
511629
511630
511631
511632
511633
511634
511635
511636
511637
511638
511639
511640
511641
511642
511643
511644
511645
511646
511647
511648
511649
511650
511651
511652
511653
511654
511655
5

DEFINITION

NOT RECORDED

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

```
/product="outer membrane protein 4"
/protein_id="CAB37072.1"
/db_xref="GI:4455887"
/db_xref="SPTREMBL:O86164"
/translation="MKTSIPWLVSSVLAFLSCHLOSILANEELLSPDDSFNENIDSGTF
TPKTSATYITGVDFYEPQKGTPLSDSCFKOTDNLITFLNGHSLTFPGFIDACTHA
GAAATTKANKLITFGSLLSDSPSTVTITGGTSSAGGVNLEIRKLVAGNES
TADGAIFKAGFLITGSDALFNNSSSTKGAATATAGARIANNTGYVRFNSIAS
TSGAIDDEGTSILNNKFLYEGNAKTTGCAICNTKASGPELIIINNNKLLIFASN
VARYSGGAHAKKLASGGSTPELRNNVSSATPKGSAISIDASGELSLSAETGNITF
VNTYTGSTGIDTAPARNAINIGSGKFTTELRAKNHTIFFYDPTISEGTSDDVLKINN
CSAGALNPQYGTITFGSTLTADDELKVDNLSKFTQPVLSGGKLLQKGVLESTG
FSQAGSLTGDIISGTTAGSITITNLGINVDSGLKQPVSTAKGASNKVLYSGK
LNLIDIEGNIYESHMFSDQLSLLKITVDADVDNDISSLIPVAPEDPNSEYFGQG
QNNYWTDTATNTKEATATWTKCFVSPERKSALCVNTLWGVFTDIRSLQQLVEIG
ATGMEHKQGFVSSWNTNLFHKTGDENRKGFRHTSGGYVIGSAHTPKDDLTFFAEHL
FARDKDCFIANNSTYGGTITFFKHSHTLQPNYLRLGRKAFSESAIEKFPRIPLAL
DVQVFSHSDNRMETHTYLSLPESEGSNECITAGIGLDLPVLSNPHPIFKTPIPM
KVMYVYQNSGFFESSDGRGFSIGRLNLNLSIPYGAKEVQDIDGSDTYDLSGFVSD
VIRNPQSTATLVMSPDMSKIRGGNLSRQAFLLRGSNNVYVNSNCELPCHYAMELRGS
SRNYNDVGTKLRF"
sig_peptide
14579..14644
/gene="omp13"
14579..16123
/gene="omp13"
14579..16123
/gene="omp13"
/codon_start=1
/transl_table=11
/protein_id="CAB37073.1"
/db_xref="GI:4455888"
/db_xref="SPTREMBL:O923D6"
/translation="WTILRNELTCSALFLALPAAQVYVYLHESDGYNGAINNKSLEPK
ITCYPEGSTYIFLDDVRISNVKHDQDAGVFINRSGNLFNMGNCNFTFHNLMTEGFG
AAISRNGDITLITNSFYLAFTSAPLLQPGQGIYSIGSVMIENSEEYTCGNYSSW
SGAAITYPLLGSKASRPSVNLGNRYLVFRDNVSOQYGGAIHNLTLTTRGPSQCFE
NNAHYHDVNSGGAIAIAPGSGISLSVKSGDILFKGNTASQDCNTIHNSIHLQSGAOF
KNLRAVSEGVYFPDIPSHSESHKITDLVINAECKEYEGTISFSGLCDDHEVCAL
NLTSLILOYDTLAGTISLSGVTIQLHSFKQEAESTLMTSPGTTLLCGSDARVONLI
ILIEDTNDVFPVRIARAEDKDALVSLKLVAFVAYSVYDFPFQFREAITPILELLUGP
alignment_scores:
Quality: 947.50 Length: 910
Ratio: 1.771 Gaps: 32
Percent Similarity: 58.791 Percent Identity: 31.648
alignment_block:
US-09-677-752-4 x CPN133034
Align seg 1/1 to: CPN133034 from: 1 to: 17280
85 ThrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerIleG1 101
||||| :||||| :||||| :||||| :|||||
3980 ACAATGCAGATGCACTATCTAATACTTAACAGGGGATGCTCAATCAC 4029
||||| :||||| :||||| :||||| :|||||
101 nAsnPheArgPhe...LeuSerPheThrAspCysSerSerLysGluSerS 117
||||| :||||| :||||| :||||| :|||||
4030 CAATGCAGGATCTCCGACAGCTCTAACCGCTTCTCTGCTTTAAGAAACTA 4079
||||| :||||| :||||| :||||| :|||||
117 erProSerIleIleHisGlnLysAsnGly...GlnLeuSerLeuArgAsn 132
||||| :||||| :||||| :||||| :|||||
4080 CTGGGAATCTTTCTTCCAGGCCAGCGCTACCAATTTCTCTACAAAT. 4129
||||| :||||| :||||| :||||| :|||||
133 .....AsnGlySe 135
||||| :||||| :||||| :||||| :|||||
4130 ATCGATGCGGAGGCAACTGTACCTTTACCAATACAGCTGCAAAATAGCT 4179
||||| :||||| :||||| :||||| :|||||
135 rMetSerPhe.....CysArgAsnH 142
||||| :||||| :||||| :||||| :|||||
4180 TCTCTCCTTTTTCAGGATTCCTATTGTCATAATAACAAACACGAAATG 4229
||||| :||||| :||||| :||||| :|||||
142 isAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSerLeuGln 158
||||| :||||| :||||| :||||| :|||||
5085 GAGAGAACTCTCGGAACGACGAAGCTGCGAGAAGCTGATAATCTCAAATCT 5134
||||| :||||| :||||| :||||| :|||||
```

20330349
2 (bases 1 to 300650)
Shirai.M.
Direct Submission
Submitted (04-JUL-2000) to the DBJ/EMBL/GenBank databases.
Mutsunori Shirai, Yamaguchi University School of Medicine,
Department of Microbiology, 1-1-1 Minamikogushi, Ube, Yamaguchi
755-8505, Japan (E-mail: mshirai@pc.cc.yamaguchi-u.ac.jp,
Tel:81-836-22-2227, Fax:81-836-22-2415)
On Sep 15, 2000 this sequence version replaced gi:6172290
gi:6172292 gi:6172294 gi:6172296 gi:6172326 gi:6172328 gi:6172330
gi:6172332 gi:6172334 gi:6172336 gi:6172338 gi:6172382 gi:6172384
gi:6172386 gi:6172388 gi:6172390 gi:6172392 gi:6172394 gi:6174666
gi:6174668 gi:6635174 gi:6635176 gi:6635178 gi:6635180 gi:8547433
gi:8547438 gi:8978640.
AB033782-AB033785, AB033800-AB033815: Submitted (25-Oct-1999)
AB038348-AB038349: Submitted (14-Feb-2000)
AB036079-AB036082: Submitted (18-Dec-2000).

FEATURES
source
1..300650
/organism="Chlamydomydia pneumoniae J138"
/strain="J138"
/db_xref="taxon:138677"
/note="synonym:Chlamydia pneumoniae (strain J138)"
144..935
/gene="CPj0267"
144..935
/gene="CPj0267"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98477.1"
/db_xref="GI:8978641"
/translation="MSINSLNKTALLNOPEPVCNLNMDPKYINQDRKTFACVTLL
VTATLMLTGVVIVLLMGSPCLSVLSTIGTSTLTGTFIIGLVKLIKLSLAMI
QVQKFOEVRVKAYEPFSLPKNDNVHKLTSCLPSLDIESPSPASTPVSKRLACSG
VAIVGLDILICAVSVVFCTGYLOIALCVGCACTGATFVCGGLATHTSLIAQGIM
YLYTYLLSALFEERNETVKDORNEINTYITTECRQKREKALLE"
957..1343
/gene="CPj0268"
957..1343
/gene="CPj0268"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98478.1"
/db_xref="GI:8978642"
/translation="MSQCQSSSTSTWEMKSVVPMKNKPTPLPSIPSEDEFTLAYEP
FVLKTPDPAQANPGTSPVNGIDNLNLLGQPNQNNANPNPTSGSNPTSLPA
PERLPTEENSQEEQSGQNNPDLIG"
complement(1495..2472)
/gene="CPj0269"
complement(1495..2472)
/gene="CPj0269"
/codon_start=1
/transl_table=11
/product="dipeptidase"
/protein_id="BAA98479.1"
/db_xref="GI:8978643"
/translation="MTIDMHCDDLSPHPCFKDPAVRCSPEQLLSGGVQRQVCAIFVP
HSRGENCDKQNSLFLSPNOYPDIGLISYEENGSSQKSLRSISINASALGD
DTAPLCTLLAKLIHLTKQGPLAYLGIVMKGNRFGCGTEAPKRSNDGKVLIDIMVEL
GVPIDLSHCSDKLAEIDLOYTADKLPNLAVIAHSRFSRVLDHRNLVDHAKETVRR
KGVIGLVRSYVGDLSGLDLEKHLVHAENLGLISVILGSDFFYANEDENFEFFNECSS
AEARPVNLQIHRIFSKGAEKSTLSRAEKLKQVIVQVNPKITDVAL"
complement(2487..3347)
/gene="ywlC"
complement(2487..3347)
/gene="ywlC"
/codon_start=1
/transl_table=11
/product="SuA5 superfamily-related protein"
/protein_id="BAA98480.1"

1773 AAGGGCAGCCAGGCCAATCTATCTATTCTATGATCCGATTGCATCTA 1724
358 luSerSerSerLysGluSerProLeuProSerSerLeuGlnAlaSerVal 374
1723 ACACCAAGAGGCTTCAGAGCTTCG..... 1698
375 ThrSerProThrProAlaThrAlaSerProLeuValIleGlnThrSerAl 391
1697 ACCATCAACCAACCGGATAGCACTCGCTTTA.....GATTA 1660
391 aAsnArgSerValIlePheSerSerGluArgLeuSer...GluGluGluL 407
1659 TTCAGGAAGCATGTATTCTCTGGGAAAGCTCTCTGCAGATGAAGCGA 1610
407 yThrProAsnLeuThrSerGlnLeuGlnProIleGluLeuLys 423
1609 AAGCTGCTGATACTTCACATCTATATTAAAGCAACCATTTGGCTCTAGCC 1560
424 SerGlyArgLeuValLeuLysAspAlaValLeuSerAlaProSerLe 440
1559 TCTGGAACCTTACCACTCAAGGAATGTCGAGCTTAGATGTCATGGTTT 1510
440 uSerGlnAspProGlnAlaLeuLeuIleMetGluAlaGlyThrSerLeuL 457
1509 CACACAGACTGAAGGCTCTACATCTCTCATGCAACCAAGCAACAGCTCA 1460
457 yThrSerSerAsp...LeuLysLeuAlaThrLeuSerIleProLeuHis 472
1459 AAGCAGATACCTAAGCTATCAGCTCTTACCAAACTTGTCTGTGATCTTCT 1410
473 SerLeuAspThrGluLysSerValThrIleHisAlaPro..... 485
1409 GCCTTAGAGGAATAGAGTGTCTCCATTGAAACAGCAGGAGCCCAACAA 1360
486AsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspG 500
1359 AACTATAACTCTAACCTCTCTCTGTTTTCAGAGATAGTAGCGG..... 1314
500 luAsnPheThrGluAsnValGluLeuLeuSerLysGluGlnAsnIle 516
1313 ..AATTTTATGAAGCCATACGATCAACCAAGCCCTTCAGCGACGCTTG 1266
517 ProLeuLeuThrLeuSerLysGluGlnSerHisLeuHisLeuProAspG 533
1265 GTGGATTCACCTGCTGCTACTGCTGTAGCGATATTATATC...GATGC 1219
533 yAsnLeuSerSer.....HisPheGlyThrGlnG 543
1218 GCCTTCACCTCTCCAGTACAACTCCAGAACCTCATACGGGTATCAGG 1169
543 lyAspTrpThrPheSerTrpLysAspSerAspGluGlyHisSerLeuIle 559
1168 GACATTGGGAAGCCACTTGGCGACACATCACTGCAAAATCAGGAAC 1119
560 AlaAsnTrpThrProLysAsnTyrValProHisProGluArgGlnSerTh 576
1118 ATGATTTGGGTAAGTACGGGCTACACCCCTAACTCTGAGCGGTAGAGGCTC 1069
576 rLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMetGlnAlaValG 593
1068 CGTAGTTCCCGATTCATTATGGGCATCTTTTACTGACATTCGCACCTAC 1019
593 lnSerMetIleAsnThrIleAlaHisGlyGlyAlaTyrLeuPheGlyThr 609
1018 AGCAGATCATGACATCTCAAGCGCAATAGTATCTATCAGCAACGAGGACTC 969
610 TrpClySerAlaValSerAsnLeuPheThrAlaHisAspSerSerGlyLy 626
968 TGGGCATCAGGAAGTGGCAATTTCTTCAT...AAGGATAAATCAGGAAC 922
626 sProIleAspAsnTrpHisHisArgSerLeuGlyTyrLeuPheGlyIleS 643
921 TAAC...CAGCATTCGCACATAAAGCTACGGCTATATTGTTGGAGAA 875

643 erThrHisSerLeuAspAspHisSerPheCysLeuAlaAlaGlyGlnLeu 659
874 GTCTCAAGATTTTCTGAAATATATCTTCTAGTGTAGCTTTCTGCCAGCTC 825
660 LeuGlyLysSerSerAspSerPheIleThrSerThrGluThrThrSerTy 676
824 TTCGGTAAAGATAAAGACCTGTTTATAGTTGAAATACCTCTCATACATA 775
676 rIleAlaThrValGlnAlaGln..... 683
774 TTTAGCGTCGTATACCTGCAACATCAGCAGATTCTCTAGGAGGACTCCCA 725
684LeuAlaThrSerLeuMetLys..... 690
724 TGCCTCATTTTGAAGTATCACCAGCATCTGCTGAAAGATATCTCTCATTT 675
691 IleSerAlaGlnAlaCystYrAsnGluSerIleHisGluLeuLysThrLy 707
674 TTCAATGCCCAGCTAAGCTACAGCTACACTAAATAATGATATGGATACTCG 625
707 sTyArgSerPheSerLysGluGlyPheGlySerTrpHisSerValAlav 724
624 CTACTTCTCTAT...CCTGAAGCTCAAGGCTCTTGGACCAATAAATCTG 578
724 alSerGlyGluValCysAlaSerIleProIle...ValSerAsnGlySer 739
577 GGCTCTAGAGCTCGGAGGATCTCTGGCTCTATATCTCCCTAAAGAAGCA 528
740 GlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGlyPheSerGl 756
527 CCGTTCTCCAGGATATTTCCCTCTTAAAGTTCCAGGCGAGCTACAG 478
756 yThrGlnAspGlyPheGluGluSerSerGlyGluIleArgSerPheSerA 773
477 CCGCCAACAAACTTTAAAGAGAGTGGCGTGAAGCCGCTGCTTTGATG 428
773 laSerSerPheArgAsnIleSerLeuProIleGlyIleThrPheGluLys 789
427 ATGAGACCTAGTCACTGCTCTCTCCCTGTCGGCATTCGGTTAGAAAAA 378
790 LysSerGlnLysThrArgThrTyrTyrPheLeuGlyAlaTyrIleGl 806
377 ATCTCCGAAGATGAAAAATAATTTCCGAGATTTCTCTAGCCTACATTGG 328
806 nAspLeuLysArgAspValGluSerGlyProValValLeuLeuLysAsnA 823
327 TGATGTGTATCGTAAAAATCCCGCTTCGCGTACTCTCTAATGTCAGTG 278
823 laValSerTrpAspAlaProMetAlaAsnLeuAspSerArgAlaTyrMet 839
277 GAGCCTCTGGACTTCGCTATGTAAAAACCTCGCACGACGAGCCCTCTTA 228
840 PheArgLeuThrAsnGlnArgAlaLeuHis...ArgLeuGlnThrLeuLe 855
227 GCAAGTCTGGAAGCCATCTGACTCTCTCCCTCATGTAGAACTCTCTGG 178
855 uAsnValSerCysValLeuArgGlyGlnSerHisSerTyrSerLeuAspL 872
177 GGAAGCTGCTTATGAGCTTCGTGCTCAGCACACATCTACAATGTAGATT 128
872 euGlyThrThrTyrArgPhe 878
127 GTGGGCTAAGATACCTCATTC 108

seq_name: gb_bai:AE001627

seq_documentation_block: 15068 bp DNA BCT 01-DEC-2000
LOCUS AE001627
DEFINITION Chlamydia pneumoniae section 43 of 103 of the complete genome.
ACCESSION AE001627 AE001363
VERSION AE001627.1 GI:4376721
KEYWORDS


```

1239 TTCAGGAACGATTGTATTTCTGGGAAAAAGCTCTCTGCAGATGAAGCGA 1288
407 yThrProAspAsnLeuThrSerGlnGlnGlnProIleGluLeuLys 423
1289 AGCTGCTGATACCTTACATCTATATTAAGCAACCATGGCTCTAGCC 1338
424 SerGlyArgLeuValLeuLysAspArgAlaValLeuSerAlaProSerLe 440
1339 TCTGGAACCTTAGCACTCAAGGAATGTGCGAGTTAGATGTCATGTGTT 1388
440 uSerGlnAspProGlnAlaLeuLeuLeuMetGluAlaGlyThrSerLeuL 457
1389 CACACAGCTGAAGGCTCTACACTCTCTCTGCAACCAAGCAACAAAGCTCA 1438
457 yThrSerSerAsp...LeuLysLeuAlaThrLeuSerIleProLeuHis 472
1439 AGCAGATACCTGAGCTATCAGTCTTACCAAACTTGTCTGTGATCTTCT 1488
473 SerLeuAspThrGluLysSerValThrIleHisAlaPro..... 485
1489 GCCTTAGAGGAATTAAGAGTGTGCTCATTAACACAGCAGGAGCAACAA 1538
486 .....AsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspG 500
1539 AACTATAACTCTAACCTCTCTCTCTGTTTTCGAAGATAGTAGCGGC.... 1584
500 LuAsnPheTyrGluAsnValGluLeuLeuSerLysGluGlnAsnAsnIle 516
1585 ..AATTTTATGAAGGCCATACGATAAACCAAGCCTTCACGCAACCTTTG 1632
517 ProLeuLeuThrLeuSerLysGluGlnSerHisLeuHisLeuProAspGI 533
1633 GTGGTATTCTACTGCTGCTACTGCTGCTAGCGATATTTATATC...GATGC 1679
533 yAsnLeuSerSer.....HisPheGlyTyrGlnG 543
1680 GCTTCTCACTTCTCCAGTACAACTCCAGAACCTCATACGGGTATCAGG 1729
543 LyAspTTrpThrPheSerTrpLysAspSerAspGluGlyHisSerLeuIle 559
1730 GACATTGGGAAGCCACTTGGCAGACACATCACTGCAAAATCAGGAAC 1779
560 AlaAsnTrpThrProLysAsnTyrValProHisProGluArgGlnSerTh 576
1780 ATGACTTGGGTAACTTCGGGCTACACCTTAATCTGAGCGTAGAGCTTC 1829
576 rLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMetGlnAlaValG 593
1830 CGTAGTCCCGATTTCATATATGGCATCTCTTACTGACATTCGCACCTAC 1879
593 InSerMetIleAsnThrIleAlaHisGlyGlyAlaTyrLeuPheGlyThr 609
1880 AGCAGATCATGACATCTCAAGCAATAGTATCTATCAGCAGCAGGAGCTC 1929
610 TrpGlySerAlaValSerAsnLeuPheTyrAlaHisAspSerSerGlyL 626
1930 TGGCATCAGGAACCTGCGAATTTCTTCCAT...AAGGATAAATCAGGAAC 1976
626 sProIleAspAsnTrpHisHisArgSerLeuGlyTyrLeuPheGlyIleS 643
1977 TAAC...CAAGCATTCGACATTAAGGCTACGGCTATATTTGGAGGAA 2023
643 erThrHisSerLeuAspHisSerPheCysLeuAlaAlaGlyGlnLeu 659
2024 GTGCTGAAGATTTTCTGAAATATCTTCAAGTGTAGCTTTCTGCCAGCTC 2073
660 LeuGlyLysSerSerAspSerPheIleThrSerThrGluThrThrSery 676
2074 TTCGTAAGATAAGACCTGTTTATAGTTGAATAACCTCTCTCATAACTA 2123
676 rIleAlaThrValGlnAlaGln..... 683
2124 TTACGCTCGCTATACCTGCAACATCGAGACTTCCTAGGAGCACTTCCCA 2173

```


114 sGluSerSerProSerIleIleHis.....GlnLysA 125
 125 snGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPheCysArgAsn 141
 2490TCTCTAGTGAATAGTGAAGAGGGCTATGAATCCG 2526
 2527 GAGGAGCATTAAACTTAGCGAATAATCCAGTATTCTGTGTGATCAGAAC 2576
 142 HisAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSerLeuG1 158
 2577 TATTCCTCCCTGAGATGTGGAGCCATCTCTGCAAGCTTTTCTCTAAC 2626
 158 nHisAsnTyLeuPheThrAlaPheGluGluAsnSerSerLysGlyAsnG 175
 2627 CGGCTCCAGCAAGAAATCAGTTCACCACTAACTCTACTCCGAAAAAAG 2676
 175 lyGlyAlaIleGlnAlaGlnThrPheSer...LeuSerArgAsnValSer 190
 2677 GTGGAGCGATTGTCTGTACGGGAATAGCTCATCTTCGGACAACCAAGGC 2726
 191 ProIleSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCy 207
 2727 ACATCAGATTTCTGGGAACACTGCTGTGAATTCGGGGGAGCAGTATA 2776
 207 sCys...SerAsnLeuIleCysSerGlyAsnValAsnProLeuPhePhe 223
 2777 TTCAGAAGCTTCATGACGATTCAGGTAAC...AACCCAGTTGCTTTTA 2823
 223 hrGlyAsnSerAlaThrAsn.....GlyGlyAlaIleCys 234
 2824 GCACAAATGCTGTTCGGGTCTATCTGATGTTGCGGTGGAGCTATCCAT 2873
 235 CysIleSerAspLeuAsnThrSerGluLysGlySerLeuLeuAlaCy 251
 2874 TGT.....ACCAAAACAGGTTACGACCCAGCCCTTACTATAAGAGA 2914
 251 sAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysGluGlyGlyA 268
 2915 TAACAAAGCTCTGATTTTTCAGGAAAACTCTTTCAGCAAAAGGTGCAG 2964
 268 laIleTyAlaLysHisMetValLeuArgTyraAsnGlyProValSerPhe 284
 2965 CGATTTACCCGATAAACTCATATTGACTTCTGGTGGCGCTACGGCATTT 3014
 285 IleAsnAsn.....SerAlaLysIleGlyAlaIleAlaI 297
 3015 ATCAATAACAAAGTTACCCATGCTACACCTAAGGGTGGAGCTATTGGTAT 3064
 297 eGlnSerGlyGlySerLeuSerIleLeuAlaGlyGluGlySerValLeuP 314
 3065 TGCTGCCAATGGAGAATGTAGTTAAACCGCTGAACATGGGGATATTACTT 3114
 314 heGlnAsnAsnSerGlnArgThrSerAspGlnGlyLeuVal...ArgAsn 329
 3115 TTGATAATAACCTGATGGCCACACAGACAACTGCTACAATAAAGAAAT 3164
 330 AlaIleTyLeuGluLysAspAlaIleLeuSerSerLeuGluAlaArgAs 346
 3165 GCATTAACATTAAGGCAATGGTAATTCGTAACCTTACGTGACGGGTC 3214
 346 nGly...AspIleLeuPhePheAspProIleValGlnGluSerSerL 362
 3215 TGGAAAGACCATTTCTTCTATGATCCTATCAGATTGAAGGTAATGCTG 3264
 362 ysGluSerProLeuProSerSerLeuGlnAlaSerValThrSerProThr 378
 3265 CTGATCTCTCACTTTGAAATAAGCTGAGGGTGATAAAGC..... 3305
 379 ProAlaThrAlaSerProLeuValIleGlnThrSerAlaAsnArgSerVa 395
 3306TATAATGGAGAAAT 3319

395 lIlePheSerSerGluArgLeuSerGluGluGlu...LysThrProAspA 411
 3320 TATTTTTTTCAGGAGAAAAGCTCACTAAGAACACAGCTCTCTTGGCGATA 3369
 411 snLeuThrSerGlnLeuGlnProIleGluLeuLysSerGlyArgLeu 427
 3370 ACCTAAAGACAAACATTTACAGCCTTATCACTTTAGCTGCTGGTGAACCTT 3419
 428 ValIleLysAspArgAlaValLeuSerAlaProSerLeuSerGlnAspPr 444
 3420 GTGTTACCGAGCGGTGCGAAGTAGAAGCAAAACACAGCTGCTGCAACACG 3469
 444 oGlnAlaLeuLeuIleMetGluAlaGlyThrSerLeuLysThr...SerS 460
 3470 AGGATCTTTGATTCTGTATGATGATGATGATGATGATGATGATGATGATG 3519
 460 erAspLeuLysLeuAlaThrLeuSerIleProLeuHisSerLeuAspThr 476
 3520 AAGATGCTACACTGACCAATCTGGCTATTATCCGAATACCTTAGATGGG 3569
 477 GluLysSerValThrIleHisAla.....ProAsnLeuSerIle 489
 3570 AAAAAATTCGCCGTAGTGCATGCGTTCGCTGGGGAAGATGTGACTTT 3619
 489 eGlnLysIlePheLeuSerAsnSerGlyAspGluAsnPheTyrgluAsn. 505
 3620 ATCAGTGCTTATTCGCGCTTATTCATCTACAGGCAAGCTTTTATGAAAC 3669
 506ValGluLeuLeuSerLys 511
 3670 ATAAGCTAAATGATACGTTAGCTTTAGGAGGAATTCACACTTTCTGGGAA 3719
 512 GluGlnAsnAsnIleProLeuLeuThrLeuSerLysGluGlnSerHisLe 528
 3720 GGTTCG.....GTGACAAACCAACCAACGCTGCTACTCATGT 3754
 528 uHisLeuProAspGlyAsnLeuSerSerHisPheGlyTyrglnGlyAsp 545
 3755 T.....GTTGGTGTCTGAAACCCACTATGTTTATCAAGGAACT 3795
 545 rPheThrPheSerTrp...LysAspSerAsp.....GluGlyHisSer 557
 3796 GGCTGTGCTGGGTCAAGATAAATAAATCTGATCTCTAAACACAAACA 3845
 558 LeuIleAlaAsnTrpThrProLysAsnTyValProHisProGluArgG1 574
 3846 GCATCTTTTACCTGGAAATAAACAGGATATGTTCCAAATCCTGAACGTCG 3895
 574 nSerThrLeuValAlaAsnThrLeuTrpAsnThrTyTrSerAspMetGlnA 591
 3896 TGCTCCGCTAGTACTCAATAGCTTTGGGATCCTTTATAGATTACGTT 3945
 591 lAlaValGlnSerMet.....IleAsnThrIleAlaHisGlyGly 603
 3946 CTATTCAGATGTCTTGGAAACGATGTTGATAGTATTCTTGAGACACGT 3995
 604 AlaTyLeuPheGlyThrTrpGlySerAlaValSerAsnLeuPheTyAl 620
 3996 CGT.....GGTTGCTGGTCTCTGGAAATTCGGGAACCTTCTTC 4031
 620 aHisAspSerSerGlyLysProIleAspAsnTrpHisHisArgSerLeuG 637
 4032 .CATAAGATCGGAATGCTCAAAATCCAAATCCCGTCATATCAGTTCCG 4080
 637 lTyTrLeuPheGlyIleSerThrHisSerLeuAspHisSerPheCys 653
 4081 GATATGTGTTAGGACCAACAAATACCTCGAGAGAGGATTCTCTTAGT 4130
 654 LeuAlaAlaGlyGlnLeuLysSerSerSerSerPheIleThrSe 670
 4131 GTGCTTTCTGCTCAGTTATTTCGAAAAGATAAGACCTACCTTGTAAACAA 4180
 670 rThrGluThrThrSerTyIleAlaThrValGlnAlaGlnLeuAlaThrS 687

1577 GG.....AATGCTTATGAAATCAGCAC...TTAGGAAAACTCAAGAC 1617

515AsnIleProLeuLeuThrLeuSerLysGluInSerH1 527

1618 TTTCATTGTGCAGCTCTCTGCTCTGGGTACTGCAACAACACACAGATGT 1667

527 sleuHisLeuProAspGlyAsnLeuSerSerHisPheGlyTyrGlnGlyA 544

1668 TCACGCGGTTCCTACAGTAGCAACTCTACGCACATATGGGTATCAAGTA 1717

544 sPrThrPheSerThrIlyasp.....SerAspGluGlyHis 556

1718 CTTGGGGAATGACTTGGTGTGATGATACCGCAAGCACTCCAAAGACTAAG 1767

557 SerLeuIleAlaAsnTrpThrProLysAsnTyrValProHisProGluAr 573

1768 ACAGCGACATTAGCTTGGACCAATACAGGCTACCTTCGGAATCCTGAGCG 1817

573 gGlnSerThrLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMetC 590

1818 TCAGAGCACTTAGTTCCTAATAGCCTTTGGGATCTTTTTCAGACATCC 1867

590 InAlaValGlnSerMetIleAsnThrIleAlaHisGlyGlyAlaTyrLeu 606

1868 AAGCAATCAAGGTGTCATAGAGAAAGTCTTGTGACTCTTTGTCAGAT 1917

607 PheGlyThrTrpGlySerAlaValSerAsnLeuPheTyrAlaHisAspSe 623

1918 CGAGGCTCTGGGCTCGGGAGTCGCCAAT...TTCATTAGATAAAGATAA 1964

623 rSerGlyLysProIleAspAsnTrpHisHisArgSerLeuGlyTyrLeuP 640

1965 GAAGAGGGAAAAA...CCCAATAACCGTCATAAATCTGGTGGATATGCTA 2011

640 heGlyIleSerThrHisSerLeuAspAspHisSerPheCysLeuAlaAla 656

2012 TCGAGGTCCAGCGCAACTTGTCTCAAAACTTAAATTAACCTTTGCTT 2061

657 GlyGlnLeuLeuGlyLysSerSerAspSerPheIleThrSerThrGluTh 673

2062 TGCCCAACTCTTTGGTAGCGATAAAGATTCTTACTGCTCAAAATCATAC 2111

673 rThrSerTyrIleAlaThrValGlnAlaGlnLeuAlaThrSerLeu... 688

2112 TGATACCTATGCAAGAGCGCTTCTATATCCACACATTACAGAAATGTAGTG 2161

688 688

2162 GGTTCATAGTGTCTCTTAGATAAATCTCTGGCTCTTGGAGTCATAAA 2211

689 ...MetLysIleSerAlaGlnAlaCysTyrAsnGluSerIleHisGluLe 704

2212 CCCCTCGTTTAAAGAGGCGAGCTCGCTTATAGCCACGTCAGTAAATGATCT 2261

704 ubysThrLysTyrArgSerPheSerLysGluGlyPheGlySerTrpHis 721

2262 GAAGACAAAGTACTTGGGTAT...CCTGAGGTGAAAGGTTCTTGGGGGA 2308

721 erValAlaValSerGlyGluValCysAlaSerIleProIleValSerAsn 737

2309 ATAATCGCTTTAACATGATGTGGGAGGTTCTCTCATCTTATCTGTA 2358

738 GlySerGlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGlyPh 754

2359 TACCTGCATGTTTGTATACCTATGCTCCATCATCACTCAAACTGAATCTGAC 2408

754 eSerGlyThrGlnAspGlyPheGluGluSerSerGlyGluIleArgSerP 771

2459 TTGATCAGCAACACCTCTCAATTTATCTTTCCCTATAGGGGTCAAGTT 2508

2409 CTATATACGTCAGGACAGCTTCTCGGAGAAAGGTACAGAAGGAAGATCTT 2458

771 heSerAlaSerSerPheArgAsnIleSerLeuProIleGlyIleThrPhe 787


```
ILCKETFTSYLRYAQCELPVTEVSQIIKSSSETDXQVLPDKVOREMEVLELRYDRV
TLNIGLKASALNPYEHVKEIILQPLLPIDKVLIVLIVSACSLNNSAAASVLEWVAIQ
LVSQYKEKELVSYSPQDLN"
/translation="tdk"
complement(5246..5866)
/sequence="tdk"
complement(5246..5866)
/codon_start=1
/transl_table=1
/product="thymidylate kinase"
/protein_id="BAA98483.1"
/db_xref="GI:8978647"
/translation="MFVIEGEGSGKSSAKALGDOLVAQDRKVLITREPQGLIGE
RLRLDILPEPHLEISRCCELEFLGSRQAHIQOEVIIPALRGYIVICERFHDSTIVYQ
GIAEGLGADFVADLCSKVGPTFPFNPVLLDIPADIGLQRIKQKRVDFEKKPLS
YHNRIREGFLSLASADPSLVLDARESLASLIDKMLHTQLGCT"
complement(5871..8375)
/sequence="gyrA_1"
complement(5871..8375)
/sequence="gyrA_1"
/codon_start=1
/transl_table=1
/product="DNA gyrase subunit A"
/protein_id="BAA98484.1"
/db_xref="GI:8978648"
/translation="MFNKEDEIIVPKNLEEMKESYLKYSMSVILSRALPDIDRGLKPS
QRRVLYAMKQLSLSPGARHKRCAKIGDTSYDHPHGESVIYPTLVRAQNMWRYPL
VGGNGFSGIDGPPAAMRYTEARLTHSAMVLMEDLDKDVIVNDETIEHPVYFP
SKFPLNLCGSSGIVAGWATNIPHNIGELIEATLLLANPQASDEILOVMPGPDFP
TGGTICGSEGRISYTHYTGKGIKVPARLHVEENEDKHRESIIITEMPYNVKSRILIE
IANVNYETLAGISDVORESDKGRVLEIKKGESSEIIINRLKFTDVQVTFGANM
LALDKLPRTMSIEMHRIEISFETPEQALALIELRLYQLTGLAEKIOKEYEELNK
LVKTRRESNGEHAERIEISFETPEQALALIELRLYQLTGLAEKIOKEYEELNK
TAYKQVGLKMPVKVYFKEQRRGGHGVTFDMKKGAGFLKAVYSFTKDYLLIFTNFGO
TISGDVYKMPVKVYFKEQRRGGHGVTFDMKKGAGFLKAVYSFTKDYLLIFTNFGO
CWLKWOLPEGERAKKPIINFLGIRPCBELAAIINIKNFNAGFLFATKRGV
KRVSDAFNSPKKRGIRALEIDEGDELIAACHIVSDEEKVMLFTHLGHAVRPPHEKVR
PMGRARGVNSKNEEDKVVSCQIVTENOSVLVDCDQFGKRLSLVDFEFTNKGCV
GVRSLINERNGNVLGAPVTDHDSILMSQGOAIRINMQDVRVYMGKSTQGVRLVHL
KEGDALVSMKLSNENDEVLSSSEECSTVSLR"
complement(8391..10808)
/sequence="gyrB_1"
complement(8391..10808)
/sequence="gyrB_1"
/codon_start=1
/transl_table=1
/product="DNA gyrase subunit B"
/protein_id="BAA98485.1"
/db_xref="GI:8978649"
/translation="MDPKENYDASAITVLEGLQAVRPRPGMYIGDTGITGUHLVYE
VVDNSIDEAMAGYCSRIDVILEDDGIVIVDNGRGTPIEHERESAKQREYSALEV
LTVLIAGGFKDQSVKSGGLHGVSCVNALSEKLVATVFKDKCYOMESRGIPTV
PLOYVSVRQCEITVYVDPKIFSTCTFDRSILMKRLRELAFNLRGITIVFEDDROV
SEKVTFFEGGIGSVFVNLNKKSESEPIYICGTRVGDGGEIEFEALQWNSGV
ELVSYANNPIPTROGTHLTGTSTAIIRVINTYIRAHNLAKNNKUALTGEDI REGITA
VLSVKNPQFQETKOKLGNDSVSAQVQVGEALTIFFEENPQIARMIVDKVFVAA
QAREAKKARETLRKSALDSARLKLIDCLEKPEKEMYIVEGDSAGGSAKQGRD
RFOATLPIRGKILNVEKARLOKIFONQIEGTIAALGCGIGADNFNLSKLRVRIII
MTDADVGSHTLLTLTFFYRHMTALINECVYIAOPPLYKYSKKKDFRYILSEKEM
SYLLMUGTNESSILFSTERELRGALESFINVLDSVESFINTELEKAIPESEFLEM
KEGIGPLYLIPATQMGGRYLDSEEREALAQEETHKRIEILYKVAVFIOYQNO
LKEYGUDISYLIPOKNEIYVIGNEDSPSCNYCYTLEEVINYLKNIKRKGIEIORYKQ
LGMNADQLWDTMNEQRTLIHVSCLKDAVEADHIFTMLMGEEVPPRREFIESHALSI
RINLNDI"
complement(10812..11159)
/sequence="Crj0276"
complement(10812..11159)
/sequence="Crj0276"
/codon_start=1
/transl_table=1
/product="Crj191 hypothetical protein"
/protein_id="BAA98486.1"
/db_xref="GI:8978650"
```

```
/translation="MFLKRRKRGSGVQNKRTASPIKHAHYLHNYLQELQK IMAARP
HDAIDAMNOVFRDKYKMSQAIGFRDHIILLVKVYNSSLYALLKOTPDNDLMSLYQVA
SHVQIREIQFLG"
alignment_scores:
  Quality: 883.00      Length: 942
  Ratio: 1.541        Gaps: 31
  Percent Similarity: 60.828  Percent Identity: 30.573
alignment_block:
  US-09-677-752-4 x AP002546/rev ..
  Align seg 1/1 to reverse of: AP002546 from: 1 to: 300650
  12 CysAlaAlaIleLeuSerSerThrAlaValLeuPheGlycInAspProLe 28
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  207773 TGTTCACACTGTTTTCGCACTGCTGAAATATATAGGCCCTCTGATAG 207724
  28 u.....GlyGluThrAlaLeuLeuThrLysAsnProAsnHisValValC 43
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  207723 CTTTGACGGGAAGTACTAACACAGCAGCCTTACTCTCTAAAATACGACTA 207674
  43 ys.....ThrPhePheGluAspCysThrMetGluSerLeu... 54
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  207673 CTGCAATAGACTATCTCTCACAGGAGATATACTCTGCAAAACCTTGGG 207624
  55 ...PheProAlaLeuCysAlaHisAlaSerGlnAsp.....AspProLe 68
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  207623 GATTTCGCGACTTTTACGAAAGGGTGTGTTTCTGACACTACGGAATCTTT 207574
  68 uTyValLeuGlyAsnSerTyCysTrpPheValSerLysLeuHisIle 85
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  207573 AAGTTTCGCGGTAAGGGGTACTCA.....CTTCTTTTAAATATTA 207530
  85 hrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerLeuGln 101
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  207529 AGTCTAGTCTGCTGAAGGC.....GCAGCACTTCTCTGTTACA 207495
  102 AsnPheArgPheLeuSerPheThrAspCysSerSerLysGlu.....Se 116
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  207494 ACTGATAAAATCTGCTCGCTAACAGGATTTTCGAGTCTTACTTTCTTAGC 207445
  116 rSerProSerIleIle.....HisGlnL 124
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  207444 GGCCCCATCATCGGTATATACACACCCCTCTAGGAAAGGTCAGTTAAAT 207395
  124 ysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPheCysArg 140
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  207394 GTGAGGGGATCTTACACTTGTATACAAATGCACTATTTTATTAACAA 207345
  141 AsnHisAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSerLe 157
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  207344 GATTACTGTAGGAAATATGCGGAGCCATTTCTACCAAGAATCTTCTTT 207295
  157 uGlnHisAsnTyrlLeuPheThrAlaPheGluGluAsnSerSerLysGly 173
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  207294 GAAAAACAGCAGCGGATCGATTCTTTTGAAGGGAATAATCGAGCGCAA 207245
  174 .....AsnGlyGlyAlaIleGlnAlaGln...ThrPheSerLeuSer 186
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  207244 CAGGAAAAAGGTCGGGCTATTTCGTCTACTGGTACTGTAGATATTACA 207195
  187 ArgAsnValSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnG 203
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  207194 AATAATACGGCTCTACCTCTCTCTCGAACAAATATTGCTGAAGTGCAGG 207145
  203 yGlyAlaIle...CysCysSerAsnLeuIleCysSerGlyAsnValAsn 219
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  207144 TGGAGCTTAATAATAGCACAGGAACTGTACAATTACAGGGAATACGCTCT 207096
  219 roLeuPhePheThrGlyAsnSerAlaThr.....AsnGlyGly 231
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  207095 ...CTTGTATTTTCTCAAAATAGTGTGACAGCCGCCAGGAATGAGCA 207048
```



```
5884 ACAGCGACATTAGCTTGGAGCAATACAGGCTACCTTCCGGAATCCTGAGCG 5835
573 gGlnSerThrLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMetG 590
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5834 TCAAGGACCTTTAGTCTTAAATAGCTTTGGGATCTTTTTCAGACATCC 5785
590 lAlaValGlnSerMetIleAsnThrIleAlaHisGlyGlyAlaTyrLeu 606
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5784 AAGCATTAAGGTGTCATAGAGAGAGTCTTTGACTCTTTGTTGATGAT 5735
607 PheGlyThrTrpGlySerAlaValSerAspLeuPheTyrAlaHisAspSe 623
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5734 CGAGGCTTGGGTCGGGAGTGGCGCAAT...TTCTTAGATAAAGATAA 5688
623 rSerGlyLysProIleAspAsnTrpHisHisArgSerLeuGlyTyrLeuP 640
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5687 GAAGGGGAAAA...CGCAATACCGTCATAAATCTGGTGATATGCTA 5641
640 heGlyIleSerThrHisSerLeuAspAspHisSerPheCysLeuAlaAla 656
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5640 TCGAGGTGAGCGCAAACTGTCTCTGAAACTTAATTAGCTTTGCGCTTT 5591
657 GlyGlnLeuLeuGlyLysSerSerAspSerPheIleThrSerThrGluTh 673
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5590 TGCCAACTCTTTGGTAGCGATAAGATTTCTTAGTCGCTAAATCATAC 5541
673 rThrSerTyrIleAlaThrValGlnAlaGlnLeuAlaThrSerLeu... 688
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5540 TGATACCTATCCAGGAGCCCTCTATATCCACACATTACAGATGTAGTG 5491
688 ..... 688
5490 GGTTCATAGTTGTCTTAGATAAACTCTCGCTCTGGAGTCATAAA 5441
689 ...MetLysIleSerAlaGlnAlaCysTyrAsnGluSerIleHisGluLe 704
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5440 CCCCTCTTTTAGAAGGCGAGCTCGCTTAGAGCCACCTCAGTAATGATCT 5391
704 uLysThrLysTyrArgSerPheSerLysGlyGlyPheGlySerTrpHis 721
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5390 GAACACAAAGTATACTCGGTAT...CCTGAGGTCAAGGTCTTTGGGGGA 5344
721 erValAlaValSerGlyGluValCysAlaSerIleProIleValSerAsn 737
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5343 ATAATGCTTTTAACATGATGTTGGAGCTTCTTCTCATCTTATCTCGAA 5294
738 GlySerGlyLeuPheSerPheSerIlePheSerLysLeuGlnGlyPh 754
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5293 TACCTGATCTTTTGATACCTATGCTCCATACATCAACTGAATCTGAC' 5244
754 eSerGlyThrGlnAspGlyPheGluGluSerSerGlyCylulleArgSerP 771
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5243 CTATATACGTGAGCAGCTCTTCGGAGAAAGGTACAGAGGAAGATCTT 5194
771 heSerAlaSerSerPheArgAsnIleSerLeuProIleGlyIleThrPhe 787
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5193 TTGATGACAGCAACCTCTTCAATTTATCTTTGCGCTATAGGGGTGAAGTT 5144
788 GluLysLysSerGlnLysThrArgThrTyrTyrPheLeuGlyAlaTy 804
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5143 GAGAAGTTCTCTGATGTTAATGACTTTTCTTATGATCTGACTTTATCCTA 5094
804 rIleGlnAspLysArgAspValGluSerGlyProValValLeuLeuL 821
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5093 TGTCTCTGATCTTATCCGCAATGATCCCAATGCACTACAGCACTTGTA 5044
821 ysAsnAlaValSerTrpAspAlaProMetAlaAsnLeuAspSerArgAla 837
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5043 TCAGCGGAGCCTCTTGGGAAACTTATGCCAATAACTTAGCACGACAGGCC 4994
838 TyrMetPheArgLeuThrAsnGlnArgAlaLeuHisArgLeu...GlnTh 853
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4993 TTGCAAGTGCCTGAGGAGCTACACTACGCTCTCTCTCTCTGTTGAAGT 4944
```

```
853 rLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSerTyrSerL 870
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4943 GTCGCGCAGTTTGTCTTGAAGTTCGTGATCCTCAGCGATTATAATG 4894
870 euAspLeuGlyThrThrTyrArgPhe 878
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4893 TAGATCTTGGGGTAAAGTCCCAATC 4868
seq_name: gb_pat1:A81835
seq_documentation_block: 2787 bp DNA PAT 21-JAN-2000
LOCUS A81835 Sequence 9 from Patent WO9858953.
DEFINITION A81835
ACCESSION A81835
VERSION A81835.1 GI:6731868
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 2787)
AUTHORS Madsen,A. and Birkelund,S.
TITLE NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE
JOURNAL Patent: WO.9858953-A.9 30-DEC-1998
MAUSEN ANNA SOFIE (DK); BIRKELUND SVEND (DK)
FEATURES
Location/Qualifiers
source 1..2787
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 811 a 583 c 598 g 795 t
ORIGIN
```

```
alignment_scores:
Quality: 857.00 Length: 865
Ratio: 1.684 Gaps: 24
Percent Similarity: 58.844 Percent Identity: 30.173
alignment_block:
US-09-677-752-4 x A81835
Align seg 1/1 to: A81835 from: 1 to: 2787
74 SerTyrCysTrpPheValSerLysLeuIleIleThrAspProLysGluAl 90
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
352 ACGTTTATAGGTTTCTTCGCTATCTTTTATTCGGTCTCCTCGAAGTTC 401
90 aLeuPheLysGlyLysAspLeuSerIleGlnAsnPheArgPheLeu 107
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
402 GATAACTACCGCAAGGAGCGCTAGC..... 429
107 erPheThrAspCysSerSerLysGluSerSerProSerIleIleHisGln 123
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
430 .....TGCTCTAGC..... 438
124 LysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPheCysAr 140
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
439 .....GGTAGCTTGAAGTTTGACAAAAATGTCAGTTTGTCTTTCAGCA 482
140 gAsnHisAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSerL 157
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
483 AAATTTTTCACGGATAATGGCGGTGCTATCACCGCAAAACTCTTTCAT 532
157 euGlnHisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerLysGly 173
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
533 TAACAGGAGCTACAACTGTCAGCTGTGTTTCTGAAATACCTCTCTCAAG 582
174 AsnGlyGlyAlaIleGln...AlaGlnThrPheSerLeuSerArgAsnVa 189
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
583 AAAGCGGAGGCATTACAGCTTCGATCCGCTTACCATTACTCGAAACCA 632
189 lSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaI 206
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```



```
749 erLysLeuGlnGlyPheSerGlyThrGlnAspGlyPheGluGluSerSer 765
|||||
2393 TGAATTGCACATTTGCTATGCACATCAGGAAGGCTTTTAAAGAACAGGGA 2442
|||||
766 GlyGluIleArgSerPheSerAlaSerSerPheArgAsnIleSerLeuPr 782
|||||
2443 ACAGAACTCGTGAATTTGGANGTAGCGCTTGTGTAATCTTGCTTACC 2492
|||||
782 oileGlyIleThrPheGluIleLysSerGln...LysThrArgThrTyrT 798
|||||
2493 TATCGGATCGGATTTGATAGGAATCAGACTGCCAAGATCAACGTACA 2542
|||||
798 yrTyrPheLeuGlyAlaTyrIleGlnAspLeuLysArgAspValGluSer 814
|||||
2543 ATCAACTCTTGGT...TATACTGTGGATCTTGTTCGTAGTAACCCGAC 2589
|||||
815 GlyProValValLeuLysAsnAlaValSerTrpAspAlaPrometal 831
|||||
2590 TGTACGCAACACTCGCAATTAGCGGTGATCTTGGAAACCTTCGGTAC 2639
|||||
831 aasnLeuAspSerArgAlaTyrMetPheArgLeuThrAsnGlnArgAlaL 848
|||||
2640 CAATTGGCAAGACAGCTTTAGTCCCTCGTGACGGGAACCATTTTGTCT 2689
|||||
848 euHis...ArgLeuGlnThrLeuLeuAsnValSerCysValLeuArgGly 863
|||||
2690 TTAACCTCAATTTGAAGCCCTTAGCCCAATTTCTTTTGAATTCGTGGG 2739
|||||
864 GlnSerHisSerTyrSerLeuAspLeuGlyThrThrTyrArgPhe 878
|||||
2740 TCATCTCGCAATACATGTAGACTTAGGAGCAAAATACCAATTC 2784
|||||
seq_name: gb_pat1:A81851

seq_documentation_block:
LOCUS A81851 3000 bp DNA PAT 21-JAN-2000
DEFINITION Sequence 25 from Patent WO9858953.
ACCESSION A81851
VERSION A81851.1 GI:6731876
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.
REFERENCE
1 (bases 1 to 3000)
AUTHORS Madsen,A. and Birkelund,S.
TITLE NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE
JOURNAL Patent: WO 9858953-A 25 30-DEC-1998.
MADSEN ANNA SORIE (DK); BIRKELUND SVEND (DK)
FEATURES
source
1..3000
Location/Qualifiers
/organism="unidentified"
/db_xref="taxon:32644"
259..>3000
/note="unnamed protein product"
/codon_start=1
/db_xref="GI:6731877"
/translation="MKSQFSLVLSSTLACFTSCSTVFATAENIGPSDFDGSNTNG
TYTPKNTTGIDYLTGLDITLONIGDSAAITKCCFSDTTEISLSPACKGYSLFLNLIK
SAGEAALSVTTDKNLSTLGFSSLTFLAEPSSVITPSGKAVKCGDGLFDNNGTILF
KQCBENGGAISTKNLSKNTSGISFEGNKSSATGKKGATCATGTDVITNTNAPT
LFNNIAEAGGAINGTNGTITGNTSLFSESVATAGNGALSGDADVTISGNOS
VFGNOAVANGAIYAKKLTLAGCGGGISFNNIVOGTTACNGCAISILAAGECSL
SAPAGDITFNGNAIVATPQTKRNSIDIGSTAKITNLRASCHSIFFDYDPTANTAA
DSTDITLNKADAGNSTDYSGISVFSEKLSDEKAVADNLSTLKQPVTLTAGNLVL
KRGVITDKGFTQAGSVINDMAGTLTKASTEVLTLGSLIPVDSLGEKKVVIASA
AKNVALSGPILLDDNOGNAYENHDLGKTQDFVQLSALGTATITDVPVPTVAPT
HYGQCTGWTGWDDTASTPKTKTATLATNTGVLNPERQGPLVPSNLWGSFSDIOA
IOGVILERSALTLCSDRGFWAAGVANFLDKDKKERRKRYKSGGYAIGAAOCTSNL
ISPAFCQLPGSKDPLVAKNHTDITYAGAFYLOHITECSGFICLLDLKPLGSSHKPLV
LEQOLAYSVDNLDKTKYTAYPEVKSQWNNAPNMLGASSHSYPEYLHCFTDYAPYI
KUNLTAYIRSDSFSEKTEGRSFDLSNLFSLPIGVKFEKFSDCNDFSYDLTLISYVD
LIRNDPKCTTALVISGASWETAYANNLRALQVRAGSHYAFSPMFEVLQGFVEVRGS
```

```
BASE COUNT 867 a 597 c 658 g 878 t
ORIGIN

alignment_scores:
Quality: 853.50 Length: 904
Ratio: 1.555 Gaps: 30
Percent Similarity: 60.730 Percent Identity: 30.752

alignment_block:
US-09-677-752-4 x A81851

Align seg 1/1 to: A81851 from: 1 to: 3000

12 CysAlaAlaIleLeuSerSerThrAlaValLeuPheGlyGlnAspProle 28
|||||
316 TGTTCCTACTGTTTTGCTGCAACTGCTGAAATAATAGGCCCTCTGATAG 365
|||||
28 u.....GlyGluThrAlaLeuThrLysAsnProAsnHisValValC 43
|||||
366 CTTTGCACGGAAGTACTAACACAGGCACCTATACTCTTAAATAACGACTA 415
|||||
43 ys.....ThrPhePheGluAspCysThrMetGluSerLeu... 54
|||||
416 CTGGAATAGACTATATCTCTGACAGGAGATATACTCTGCAAAACCTTGGG 465
|||||
55 ...PheProAlaLeuCysAlaHisAlaSerGlnAsp.....AspProLe 68
|||||
466 GATTCCGGCAGCTTTAACGAGGGTGTGTTTTCTGACACTACGGAATCTTT 515
|||||
68 uTyrValLeuGlyAsnSerTyrCysTrpPheValSerLysLeuHisIleT 85
|||||
516 AAGCTTTGCCGGTAAGGGTACTCA.....CTTTCTTTTAAATATTA 559
|||||
85 hrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerIleGln 101
|||||
560 AGTCTAGTGTGGAAGGC.....GCAGCACATTCTGTTTACA 594
|||||
102 AsnPheArgPheLeuSerPheThrAspCysSerSerLysGlu.....Se 116
|||||
595 ACTGATAAAATCTGTCGTAACAGGATTTTCGAGTCTTACTTTCTTAGC 644
|||||
116 rSerProSerIleIle.....HisGlnL 124
|||||
645 GGCCCATCATCGCTAATCACACCCCTCAGGAAAGGTGCAGTTAAAT 694
|||||
124 ysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPheCysArg 140
|||||
695 GTGAGGGGATCTTACATTTGATACCAATGGAACATTTTATTAAACAA 744
|||||
141 AsnHisAlaGluGlySerGlyAlaIleSerAlaAspAlaPheSerLe 157
|||||
745 GATTACTGTGAGGAAATGCGGAGCCCAITTTACCAAGATCTTTCTTT 794
|||||
157 uGlnHisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerLysGly. 173
|||||
795 CAAAACAGCAGCGGATCGATTCTTTTGAAGGGAATAAATCGAGCGCAA 844
|||||
174 .....AsnGlyGlyAlaIleGlnAlaGln...ThrPheSerLeuSer 186
|||||
845 CAGGAAAAAAGTGGGGCTTTTGTGCTACTGCTAGTACTGTAGATTTACA 894
|||||
187 ArgAsnValSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnG 203
|||||
895 AATAACGGCTCTTACCCTCTTCTCGAACCAATATTGCTGAAGCTGCGAG 944
|||||
203 yGlyAlaIle...CysCysSerAsnLeuIleCysSerGlyAsnValAsnP 219
|||||
945 TGGACCTATAATACCACAGCAAACTGTACAATTACAGGGAATAGCTCT. 993
|||||
219 roLeuPhePheThrGlyAsnSerAlaThr.....AsnGlyGly 231
|||||
```

```
788 GluLysLysSerGlnLysThrArgThrTyrTyrTyrPheLeuGlyAlaTy 804
||||| |||::: :::: ||| :::: |||
2767 GACAAGTCTCTCATGTAATGACTTTCTTCTATGACTGACTTATCCCTA 2816
||||| |||::: :::: ||| :::: |||
804 rleGlnAspLeuLysArgAspValGluSerGlyProValValLeuLeuL 821
||||| |||::: :::: ||| :::: |||
2817 TGTTCTGATCTTATCCGAATGATCCCAATGACACTACAGCAGCTGTA 2866
||||| |||::: :::: ||| :::: |||
821 ysaAsnAlaValserTTPAspAlaProMetAlaAsnLeuAspSerArgAla 837
||||| |||::: :::: ||| :::: |||
2867 TCAGCGGAGCTCTTGGAACCTTATGCCAATAACTTAGCACCACAGGCC 2916
||||| |||::: :::: ||| :::: |||
838 TyrMetPheArg 841
::: |||
2917 TTGCAAGTGGCT 2928

seq_name: gb_ba2:CPOMP54

seq_documentation_block:
LOCUS CPOMP54 6030 bp DNA BCT 22-JAN-1999
DEFINITION Chlamydia pneumoniae omp5 and omp4 genes.
ACCESSION AJ001311
VERSION AJ001311.1 GI:3255934
KEYWORDS omp4 gene; omp5 gene; outer membrane protein 4; outer membrane
protein 5.
SOURCE
ORGANISM Chlamydia pneumoniae.
Chlamydia pneumoniae
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE
AUTHORS Knudsen,K., Madsen,A.S., Mygind,P., Christiansen,G. and
Birkelund,S.
TITLE Identification of two novel genes encoding 97- to 99-kilodalton
outer membrane proteins of Chlamydia pneumoniae
JOURNAL Infect. Immun. 67 (1), 375-383 (1999)
MEDLINE 99081766
FEATURES
source
1..6030
location/Qualifiers
/organism="Chlamydia pneumoniae"
/isolate="CDC/CWL-029/VR-1310"
/db_xref="taxon:83558"
/clone="pEX2-18"
/clone="B8F3"
/clone_lib="pbluescript"
/dev_stage="elementary body"
complement(1..2804)
/gene="omp5"
complement(<1..2742)
/gene="omp5"
/feature="putative lipoprotein"
/codon_start=1
/transl_table=11
/product="outer membrane protein 5"
/protein_id="CAA04671.1"
/db_xref="GI:3255935"
/db_xref="SPTREMBL:O86163"
/translation="MKSQFSLVLSSTLACFTSCSTVFAAENIGPSDSFDGSTNTG
TYPKNTTIDYTLTGDIQLNLGDSALTKGCFSDTTESLSFAGKGYLSFLNKS
SAEGAALSVTTDKNLGFSLSLTELAPSSVITTPSGKAVKCGDLTFDNGTILF
DQCEENGALSTNLSKSTGTSISFEGNKSATGKGAICATGVTDTNNRPT
LFSNNIAEAGGAINSTGNTISLVSFSENSVATAGGALSGDADVTISQNS
VTFSGDVTGANGAIYAKKLLASGGGGISFNNIVQGTAGNGGALSILAAGCSL
SAEAGDTFNGNAIVATTPTTKRNSIDIGSTAKITNLRAISGHISFIYDPTANTAA
DSTDTLNLNKADAGNSTDYSGLVFSGEKLSDEEAKVADNLTSTLKQPVTLTAGNLVL
KRGVTLDKGFTQAGSSVINDAGTTLKASTEEVTLTGLSIPVDSISGEGKKVIAASA
AKSNVLDSPKLLLNQGNAYENHDLGKTQDFSVQLSALGTATTTDVPVPTVATPT
HYGQGTGWTWDDTASTPTKTKTATLAWNTGTLNPERQGLVPNSLWGSFSDIOA
```

```
557 SerLeuIleAlaAsnTrpThrProLysAsnTyrValProHisProGluAr 573
   ::::::::::::::::::::
975 ACAGCGACATTAGCTTGGACCAATACAGGCTACTTCCGAATCTGAGCG 926
   ::::::::::::::::::::
573 gGlnSerThrLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMetG 590
   ::::::::::::::::::::
925 TCAGGACCTTTAGTTCTCTATAGACCTTTGGGATCTTTTCAGACATCC 876
   ::::::::::::::::::::
590 InAlaValGlnSerMetIleAsnThrIleAlaHisGlyGlyAlaTyrLeu 606
   ::::::::::::::::::::
875 AAGCGATTCAAGGTCTCATAGAGAGAGTCTTTGACTCTTTGTTTCAGAT 826
   ::::::::::::::::::::
607 PheGlyThrTrpGlySerAlaValSerAsnLeuPheTyrAlaHisAspSe 623
   ::::::::::::::::::::
825 CGAGGCTCTCGGCTCGGAGTGCAGCAAT...TTCTTAGATAAAGATAA 779
   ::::::::::::::::::::
623 rSerGlyLysProIleAspAsnTrpHisHisArgSerLeuGlyTyrLeuP 640
   ::::::::::::::::::::
778 GAAGAGGGGAAAA...CCCAATACCGCTATAAATCTGGTGGATATGCTA 732
   ::::::::::::::::::::
640 heGlyIleSerThrHisSerLeuAspHisSerPheCysLeuAlaAla 656
   ::::::::::::::::::::
731 TCGGAGGTGCGAGCCCAACTTGTCTCAAACTTAATAGCTTTGCCTTT 682
   ::::::::::::::::::::
657 GlyGlnLeuLeuGlyLysSerSerAspSerPheIleThrSerThrGluTh 673
   ::::::::::::::::::::
681 TGCCAACCTCTTTGGTAGCGATAAAGATTCTTACTCGCTAAAAATCATAC 632
   ::::::::::::::::::::
673 rThrSerTyrIleAlaThrValGlnAlaGlnLeuAlaThrSerLeu... 688
   ::::::::::::::::::::
631 TGATACCTATGACGAGGCGCTTCTATATCCACACATTACAGAATGTAGTG 582
   ::::::::::::::::::::
688 .....
581 GGTTCATAGTGTGCTCTTAGATAAACTTCTCGCTCTGGAGTCATAAA 532
   ::::::::::::::::::::
689 ...MetLysIleSerAlaGlnAlaCysTyrAsnGluSerIleHisGluLe 704
   ::::::::::::::::::::
531 CCCCTCGCTTTTAGAAGGCGAGCTCGCTTATAGCCACGCTCACTAATGATCT 482
   ::::::::::::::::::::
704 uLysThrLysTyrArgSerPheSerLysGlyGlyPheGlySerThrPhis 721
   ::::::::::::::::::::
481 GAAGCAAAAGTATACTCGGTAT...CTGAGGTGAAGGTCTTTGGGGGA 435
   ::::::::::::::::::::
721 erValAlaValSerGlyGluValCysAlaSerIleProIleValSerAsn 737
   ::::::::::::::::::::
434 ATAATGCTTTTAACATGATGTGGAGAGCTTCTCTCATCTTATCCCGAA 385
   ::::::::::::::::::::
738 GlySerGlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGlyPh 754
   ::::::::::::::::::::
384 TACCTGCTGTTTGTATACCTATGCTCCATACATCAAACTGAATCTGAC 335
   ::::::::::::::::::::
754 eSerGlyThrGlnAspGlyPheGluGluSerSerGlyCluIleArgSerp 771
   ::::::::::::::::::::
334 CTATATAGCTCAGGACAGCTTCTCGGAGAAGGTACAGAAGGAAGTCTT 285
   ::::::::::::::::::::
771 heSerAlaSerSerPheArgAsnIleSerLeuProIleGlyIleThrPhe 787
   ::::::::::::::::::::
284 TTGATGACAGCAACCTTCAATTTATCTTTGCTATAGGCGTGAAGTTT 235
   ::::::::::::::::::::
788 GluLysLysSerGlnLysThrArgThrTyrTyrPheLeuGlyAlaTyr 804
   ::::::::::::::::::::
234 GAGAAGTCTCTGATGTGAATGACTTTTCTTATGATCTGACTTTTCTCTA 185
   ::::::::::::::::::::
804 rIleGlnAspLeuLysArgspValGluSerGlyProValValLeuLeuL 821
   ::::::::::::::::::::
184 TGTTCCTGATCTTATCCGCAATGATGCCAAATGCACTACAGCACTGTAA 135
   ::::::::::::::::::::
821 ysAsnAlaValSerTrpAspAlaProMetAlaAsnLeuAspSerArgAla 837
   ::::::::::::::::::::
134 TCAGCGAGGCGCTCTTGGAAACTTATGCCAATAACTTAGCACGACAGGCC 85
   ::::::::::::::::::::
838 TyrMetPheArg 841
```

seq_name: gb_bal:AE001587

seq_documentation_block:

LOCUS AE001587 16448 bp DNA BCT 01-DEC-2000

DEFINITION Chlamydia pneumoniae section 3 of 103 of the complete genome.

ACCESSION AE001587 AE001363

VERSION AE001587.1 GI:4376271

KEYWORDS

SOURCE Chlamydia pneumoniae CWL029.

ORGANISM Chlamydia pneumoniae CWL029

REFERENCE 1 (bases 1 to 16448)

AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W., Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.

TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis

JOURNAL Nat. Genet. 21 (4), 385-389 (1999)

MEDLINE 99206606

REFERENCE 2 (bases 1 to 16448)

AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.

TITLE Direct Submission

JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA

FEATURES

Source

1..16448

/organism="Chlamydia pneumoniae CWL029"

/strain="CWL029"

/db_xref="taxon:115713"

172..2957

/note="Polymorphic Outer Membrane Protein G Family; contains frame shift"

172..705

/gene="pmp_3.1"

/note="Cpn0014"

615..2957

/gene="pmp_3.2"

/note="Cpn0015"

3196..4971

/gene="pmp_4.1"

/note="Cpn0016"

3196..5953

/note="Polymorphic Outer Membrane Protein G Family; contains frame shift"

4874..5953

/gene="pmp_4.2"

/note="Cpn0017"

6302..7786

/gene="pmp_5.1"

/note="Cpn0018"

6302..9139

/note="Polymorphic Outer Membrane Protein G Family; contains frame shift"

7787..9139

/gene="pmp_5.2"

/note="Cpn0019"

complement(9380..11467)

/gene="Cpn0020"

complement(9380..11467)

/gene="Cpn0020"

/codon_start=1

/transl_table=11

/product="Predicted OMP [leader peptide: outer membrane]"

/protein_id="AAD18173.1"

/db_xref="GI:4376272"

/translation="MKRCFLFLASFLVMGSSADALTHQEA VKKNSYLSHFKSVSGIV

TIEDGVLAHNLRITQANKYVYVENTGQSLKLVAGNVNRYRAKTLVCDLYEYEDT

SLSLITNGRFAMPWFLGSMITLTPETIVIRKGYISTEGPKKDLCLGDLVEYSSD

DCSJKTKTLRVCRPIPLFLPFPSPIMPPIKPPINFRGGTGGTGLSGYLSYSPISR

KHFSSTFFLDSEFKHGVGMGNLHCSQKQPVENVNKNKSYAHRLAIDNAEDRYRL

ENRCCVREQEDYKWEKREAEFRERGNKILSPREELESSLEQFDHGLKNFSEKLMEL
EGHLLKQKENTAEVKNKILSDAESRLIEIVFEDVKEMPCRIEEIEKTLRMAELPLPPT
KKAFAKACQVNSCAEMLEKVKPKESLAYVTSKERLVSLDEDLRRAYTECQKRFOG
DSGLSEVRACREGLRIQRFETQGLDLVEKELLVSSRIINTECDVCVCKKEAPP
GKKYAOYDDE1YVRVQSRWMTMSERLREGVQACNMLKACISEEDKVLKEEYEWLY
REERKNEKRLVGTGKIIVATOORIQEOPQSDIVESSNEKVSIMDKARLFENREDHS"
12262..12987

gene

/gene="hb6"

CDS

12262..12987

/gene="hb6"

/note="hypothetical"

/codon_start=1

/transl_table=11

/product="HB6 protein"

/protein_id="CAB37081.1"

/db_xref="GI:4455897"

CGATVLSGVSGLSALVCGLGISTISLVGVFLVGLILLRKRRELTLEOLEAKOIAE
TFABELENIYIGSTKSEKIEGSRISOOGFLNRATOKILDLLESSLSITSFERDL
RQLEDEKIELLSGERLLEFTAANLFKQGRDVLNLGNLADIRAYMGNPNKVKAVVIE
KAKAVHFEIVLTTMARELEFF"
13014..14585

gene

/gene="hb7"

CDS

13014..14585

/gene="hb7"

/note="hypothetical"

/codon_start=1

/transl_table=11

/product="HB7 protein"

/protein_id="CAB37082.1"

/db_xref="GI:4455898"

/translation="MYGNERLLERLLYNSVOKSYADRLFSYKTKMVHDTPLPWEE
DKERCAFAKFLQCKILLDYKISFIWLNENDEINLNDPWSGINTVTRKVPQEV
DSEWNHKKIQLKEDNYFKLIESSEKTEANKKLLSDLVDRLEDAKTFKFLKQEE
VETRKDLARYGCTVDPKQTEAKKVELEASLETFLDSIESLVOCLEDDQDTYWE
QDVDLARTOLEEQDTEAKREAAEDLSRLERLKKSKTMDRAKHIEAEDSITW
WTSQEMDKARKLIKLEDTISVLPEDILETCLSEUEELPPLTIRELLTSYKLFKI
CSETLLKNNKRYNIVYQEVQNLQGFLOGISORFKKQDDFANLEQVQALOKK
RLRELTQNFKEDEFAAKAKDLYTRSTAEQKNFDPVCMELFRRYHEEVNKP
LLELMYNCAFSYRAKAKKLCSLRDLDEKELQKEIKKEFYQKKQORHADRSHRTTYQK
LRIAEELALELKKKI"
14934..17479

gene

/gene="omp7"

CDS

14934..17479

/gene="omp7"

/note="putative"

/codon_start=1

/transl_table=11

/product="outer membrane protein 7"

/protein_id="CAB37083.1"

/db_xref="GI:4455899"

/translation="MKLIPRLFLILSLVPTLSMSNLIGAATTSELSASNSFDGTTSTTS
FSSKTSSTADGTNIVFKDSVVIENVPKTGETQSTSCFKNDAAAGDLNPLGGGFSFTFS

alignment_scores:

Quality: 852.00 Length: 812
Ratio: 1.635 Gaps: 30
Percent Similarity: 64.163 Percent Identity: 31.158

alignment_block:

US-09-677-752-4 x CPN133035

Align seg 1/1 to: CPN133035 from: 1 to: 26920

120 lletleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMe 136

21233 CTGGTGTGGAGAGTATAGGAAATCTTGAATCTTAATGAAATGCCTCTAG 21282

136 tserPheCysArgAsnHisAlaGluGlySerGlyGlyAlaIleSerAla 153

21283 TGAA.....AATGGGGAGCCATCAATACGA 21308

153 spAlaPheSerLeuGlnIleAsnTyrLeuPheThrAlaPheGluGlnAsn 169

21309 AGACNTTGTCTTTGACTGGGAGTACCGGTTTGTACGGTCTCTTGGCAAT 21358
170 SerSerLysGlyAsnGlyGlyAlaIleGlnAlaGlnThrPheSer...Le 185
185 uSerArgAsnValSerProIleSerPheAlaArgAsnArgAlaAspLeuA 202
21409 TTCTGAGAATGCAGGAATCTTGAGCTTCGGAAACAACAGTCGCACACAT 21458
202 snGlyGlyAlaIleCysCys...SerAsnLeuIleCysSerGlyAsnVal 217
21459 CAGGAGCGCGATCTCTGCTGAAGGGAACCTTGTGATCTCCAAATAACCAA 21508
218 AsnProLeuPhePheThrGly...AsnSerAlaThrAsnGlyGlyAlaI 233
21509 AAT...ATCTTTTTCGATGGCTGCAAGCAACTACAAATGGCGGAGCTAT 21555
233 eCysCysIleSerAspLeuAsnThrSerGluLysGlySerLeuSerLeuA 250
21556 TGATTCT...AACAAAGCAGGGCGCAACCCAGACCCTCTTTCGACTCTT 21602
250 laCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysGluLysGly 266
21603 CAGGAATGAGAGCTGCTATTTCTGAATAACACAGCAGGAAATAGTGA 21652
267 GlyAlaIleTyrAlaLysHisMetValLeuArgTyr...AsnGlyProva 282
21653 GGTGCGATTATACCAAAATTTGGTGTATCTCCTCAGGAGGAGGAGT 21702
282 lSerPheIleAsnAsnSerAla.....LysIleGlyGlyAlaI 295
21703 GTTATTCTTCAACAACAGCTCGAATGCTACTCTCTAAAGAGAGGGCAA 21752
295 leAlaIleGlnSerGlyGlySerLeuSerIleLeuAlaGlyGluGlySer 311
21753 TTGCGATTCTAGATTCTGGAGAGATAGCATTTCTCGAGATCTCGGCAAT 21802
312 valLeuPheGlnAsnAsnSerGlnArgThr.....SerAspGlnGlyLe 326
21803 ATCATTTTCGAGGGCAATACTACGAGCACTACAGGAAGCTCTCGAGTGT 21852
326 uValArgAsnAlaIleTyrLeuGluLysAspAlaIleLeuSerSerLeuG 343
21853 GACCAGAAATGCTATAGATCTTCGATCGAATGCAAAATTTTAAATCTCC 21902
343 luAlaArgAsnGlyAsp...IleLeuPhePheAspProIleValGlnGlu 358
21903 GACGAGCTCGGGAAATAAAGTTATTTCTATGATCTCTATC..... 21943
359 SerSerSerLysGluSerProLeuProSerSerLeuGlnAlaSerValTh 375
21943 21943
375 rSerProThrProAlaThrAlaSerProLeuValIleGlnThrSerAla 392
21944ACGAGCTCAGGAGCTACTGATAAGCTCTCTTTGA 21977
392 snArgSer.....ValIlePhe 397
21978 ATAAAGTCAGCGAGGATCTGGAAATACCTATGAAGGCTACATCGTTTC 22027
398 SerSerGluArgLeuSerGluGlu...LysThrProAspAsnLeuTh 413
22028 TCTGGAGAGAACTCTCAGAGAGGAACTTAAGAAACCTGACATCTGAA 22077
413 rSerGlnLeuGlnProIleGluLeuLysSerGlyArgLeuValLeuL 430
22078 GTCTACATTTACACAGGCTGTAGAGCTTGCTGCAGAGTCCCTTAGTATGA 22127
430 ysAspArgAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAla 446
..... 446


```
/gene="omp5"
complement(1..2742)
/gene="omp5"
/feature="outer membrane protein 5"
/codon_start=1
/transl_table=11
/product="outer membrane protein 5"
/protein_id="CAA04671.1"
/db_xref="GI:3255935"
/db_xref="SPTREMBL:O86163"
/translation="MKSQFSLVLSLTACFTSCSTVFPAATAENIGPDSFDGSGTNG
TTPKNTGIDYITLGDITLQNGSALTKGCSDDTESLSPAGKGYSLFLNKS
SAGAAVSTTDKNLSLGFSSLTFLAAPSIVITPSGKGAVKCGDLDFFDNGNITL
KQYCEGNGAISTKNLSKSTGSPKSSATKGGKGAICATGTVDTNNTPT
VFSNIAAAGAGINSTGNTGTSVFSNSVTATAGNGALSGDADVTISQNS
LTFSGNQAVGAGIYAKLTLASGGGGISFNINVOGTTAGNGAISIILAAAGCSL
SAEAGDITFNGAIVATTPOTTKRNSIDIGSTAKITNLRASGHSIFFFDPTANTAA
DSTDILNKNKADAGNSTDYSGSIVFSGEKLSEDAKADNLSTLQKQPVTLTANLVL
KRGVLTDRKFTQAGSSVIMDAGITLKAETEEVTLGLSFVDSLGEKKVYIAASA
ASKYNALSGPILLDNQGNAYENHDLGKTQDFVQSLGATGATTIDVPAVPTVAPT
HYGQGTQMTWDDTASTPKTKTATLAWNTCYLPNEROGPLVPSNLWGSFSDIOA
IQGVIFERSALTICSORGFWAAGVANFLDKDKGKKRYRHKSGGVAIGGAOATCSNENL
ISFAPCLFGSKDDELVAKNHTDTVAGAFYIOHITECSCFICGLLDKLPKGSWHPVLV
LGQALYSHVSNDLTKYTAPEVKGSGNNAFNMNLGSSSHSYPEYLHCFTDYAPYI
KLNTYIRDSFSEKTEGRSFDSDNLFNLSLPVGVKFEKSDCNDFSVDLTSVYPO
LIRNPKCTTALVISGAWETYNANLQALQVRAAGSHYAFSPMEFVLQGFVEVRGS"
complement(1..2697)
/gene="omp5"
/product="outer membrane protein 5"
complement(2698..2742)
/gene="omp5"
complement(2747..2752)
/gene="omp5"
/feature="putative"
complement(2777..2782)
/gene="omp5"
/feature="putative"
complement(2799..2804)
/gene="omp5"
2863..5726
/gene="omp4"
2863..2868
/gene="putative"
2896..2902
/gene="omp4"
/feature="putative"
2905..2955
/gene="omp4"
2905..5691
/gene="omp4"
/feature="putative lipoprotein"
/codon_start=1
/transl_table=11
/product="outer membrane protein 4"
/protein_id="CAA04672.1"
/db_xref="GI:3255936"
/db_xref="SPTREMBL:O86164"
/translation="MKTSLPVLVSVLAFSCHLQSLANEELSPDPSFNGNIDSGTF
TPKTSATTLGDDVFYEPFGKTPSLSDSCFKQTTDNLTLFLNGHSLTEFGFDAGTHA
GAAATANKLNTLFGSLFSLSDSPSTVTQGTLSGAGVNLNIRKLVVAGNFS
TAGDKAIGKASFLTGTSGDALPSSNSSTKGAICATTAGARIANNVTGYVRLSNLAS
TSGCAIDDEGTSILSNKFLYFENAKTTGGAICATKAGSPLELIISNNKTLIPASN
VAETSGAIAKKALSLSSGFTFELNNVSSATPKGAISIDASGELSLSAETGNITF
VRNLITFGTDPKRNAINIGSKFTTELRAKNHTIFFYDITSEGSSVDLKNIN
GSAGALNPGQTTILFSGETLTADLKVADNLKSSFTQPSVLSGKLLQLQGVLETS
FSGAELGMSGHTLSTAGTITINVDLSLGLKOPVSLTAKGASNKVIVSGK
NLNLDIEGNIYESHMTSHQDLFSLLKITVDADVNTVDLSLIPVPAEDNPSYVQFG
QKNNVNTDPTATNKATATWTKTGFVPSPERKSALVCNTLWCVFTDIRSLQOLVEIG
ATGMEHKQFWNSMNTNFLHKTGDENRKGFRHTSGGYVIGGSAHTPKDDLTFATFCHL
FARDKCFTHAANSRTYGGTFFKHSHTLQPNYLRGRKAFSESAIERFPREPLAL
DVQVSFSHSDNRNETHYTSLSPESEGSWSNECIAGGIGLDLPFLVLSNPHPLFKTFIPOM
```

```
mat_peptide
sig_peptide
RBS
-10_signal
-35_signal
gene
-10_signal
RBS
sig_peptide
CDS
KVMYVYVSONSEFFSSDGRGFSIGRLNLNLSPVCAKFKVQGDIGDSYTYDLSGFEVSD
VYRNPOQSTATLVMSPDSPKIRGGLNSRQALLRGSNNYVYNSNCELFHYAMELRGS
SRNYNDVGTGTLRF"
2956..5688
/gene="omp4"
/product="outer membrane protein 4"
2958..5726
/gene="omp4"
/feature="putative"
1760 a 1341 c 1195 g 1734 t
BASE COUNT
ORIGIN
alignment_scores
Quality: 849.50 Length: 959
Ratio: 1.467 Gaps: 32
Percent Similarity: 60.375 Percent Identity: 27.737
alignment_block:
US-09-677-752-4 x CPOMP54
Align seg 1/1 to: CPOMP54 from: 1 to: 6030
7 AsnPheCysCysLeuCySAAlaAlaLeuSerThrAlaValLeuPh 23
2984 AACTTTATACCTGATGATGCTTTTAAATGGAATATCG..... 3022
23 eGlyGlnAspProLeuGlyGluThrAlaLeuLeuThrLysAsnProAsnH 40
3023 .....ATTGAGGAACGTTTACTCCAAATCTTCAGCCACAC 3059
40 isValValCysThrPhePheGluAspCysThr...MetGluSerLeuPhe 55
3060 ATATTCTCTAACAGGAGATGCTTCTTTTACGAGCCTGGAAGGACACTC 3109
56 ProAlaLeuCySAAlaHISAlaSerGlnAspAspProLeu.TyrValLeuG 72
3110 CCTTATCTGACATGTTGTTTAAAGCAACCCAGGACATCTTACCTTCTTG 3159
72 LysAsnSer.....TyrCysTrpPheValSer 80
3160 GGAACGCTCATAGCTTAACGTTTGGCTTTATAGATGCTGCACCTCATGC 3209
81 LysLeuHisIleThrAspProLysGluAlaLeuPheLysGluLys.GlyA 97
3210 AGGTCTCTGC.....TGCATCTACAACAGCAATAAGA 3241
97 spLeuSerIleGlnAsnPheArgPheLeuSerPheThrAspCysSerSer 113
3242 ATCTTACCTCTCAGGGTTCCTTACTGAGTGT..... 3276
114 LysGluSerSerProSer...IleIleHisGlnLysAsnGlyGlnLeuSe 129
3277 ...GATTCCTCTCTAGCACACGCTTACTACAGCTCAGGAGACGCTTTC 3323
129 f.....LeuArgAsnAsnGlySerMetSerPheCysA 140
3324 CTCACGAGGAGCGCTAAATTTAGAAATATTTCGTAACCTTGTAGTTGCTG 3373
140 rGAsnHisAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSer 156
3374 GGATTTTTCTTACTACGATGCTGAGCTATCAAGAGGAGCGCTTTCCTT 3423
157 LeuGlnHisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerLysG 173
3424 TTAACCTGACACTCTGGAGATGCTCTTTTGTAGTACACTCTTCAATCAC 3473
173 yAsnGlyGlyAlaIleGlnAlaGlnThrPheSer...LeuSerArgAsnV 189
3474 AAAGGAGGAGCAATTTGCTACTACAGCGCTCCCATAGCAATAACA 3523
189 alSerProLysSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAla 205
.....
```



```

688 LeuMetLysIle..... 691
6960 TATATCTCTCTCTGCAAACTACGGCTGTTCCTGCTTATGTTCC 6911
692 .....SerAlaGlnAlaCysTyrAsnGluSerI 701
6910 TACAGAGATTCCTGTTCTCTTTTTCAGAAACCTTAGCTACACCCATACGG 6961
701 leHisGluLeuLysThrLysTyrArgSerPheSerLysGluGlyPheGly 717
6860 ATACGATCTGAAACCAAGTATACACATATCTACTCTTAA...GGA 6814
718 SerTrpHisSerValAlaValSerGlyGluValCysAlaSerIleProII 734
6813 AGCTGGGGAATGATAGTTTCGCTTTAGAAATTCGGTGAAGAGCTCCGAT 6764
734 eValSerAsnGlySerClyLeuPheSerPheSerIlePheSerLysL 751
6763 TTGCTTAGATGAAGTGCTCTATTGAGCAGTACATGCCCTTCATGAAT 6714
751 euGlnGlyPheSerGlyThrGlnAspGlyPheGluGluSerSerGlyGlu 767
6713 TGCAGTTTCTGTCATGCATCAGCAAGGTTTAAAGAACAGGCAACAGAA 6664
768 IleArgSerPheSerAlaSerPheArgAsnIleSerLeuProIIeGI 784
6663 GCTCGTGAATTTGGAAGTAGCCGCTTGTGNAATCTTGCCCTTACCTACGG 6614
784 ylleThrPheGluLysSerGln...LysThrArgThrTyrTyrP 800
6613 GATCCGATTGTAAGGAATCAGACTGCCAAGATGCAACGTACAATCTAA 6564
800 heLeuGlyAlaTyrIleGlnAspLysArgAspValGluSerGlyPro 816
6563 CTCTTGGT...TATAGTGTGATCTTGTTCGTAGTAACTCCGACTGACG 6517
817 vaIValLeuLeuLysAsnAlaValSerTrpAspAlaProMetAlaAsnLe 833
6516 ACAACACTGCGAATTCAGCGTGATCTTCTGGAAACCTTTCGGTACGATTT 6467
833 uAspSerArgAlaTyrMetPheArgIleuThrAsnGlnArgAlaLeuHis 849
6466 GCGAAGCAGACGCTTGTCTCTGCGAGGACCATTTTGTCTTAACT 6417
850 ..ArgLeuGlnThrLeuLeuAsnValSerCysValLeuArgGlyGlnSer 865
6416 CAATTTTGAAGCTTTAGCCAAATTTCTTTTGAATTCGCTGGGTCACT 6367
866 HisSerTyrSerLeuAspLeuGlyThrThrTyrArgPhe 878
6366 CGCAATTACAACTAGACTTAGGAGCAAAATACCAATTC 6328

seq_name: gb_ba2:AP02545

seq_documentation_block:
LOCUS AP02545 299650 bp DNA BCT 08-DEC-2000
DEFINITION Chlamydomophila pneumoniae J138 genomic DNA, complete sequence,
section 1/4.
ACCESSION AP02545 AB033780 AB033781 AB033792 AB033793 AB033794 AB033795
AB033796 AB033798 AB033799 AB036071 AB036072 AB036073
AB036074 AB036075 AB036076 AB036077 AB036078 AB036079 AB036080
AB036081 AB036082
VERSION AP02545.2 GI:9956082
KEYWORDS
SOURCE
ORGANISM Chlamydomophila pneumoniae J138 (strain:J138) DNA.
REFERENCE 1 (sites)
AUTHORS Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,
Shiba,T., Ishii,K., Hattori,M., Kihara,S., and Nakazawa,T.
TITLE Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CWL029 from USA

```

```

JOURNAL MEDLINE
REFERENCE 2 (bases 1 to 299650)
AUTHORS Shirai,M.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases.
Mutsunori Shirai, Yamaguchi University School of Medicine,
Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi
755-8505, Japan (E-mail:mshirai@po.cc.yamaguchi-u.ac.jp,
Tel:81-836-22-2227, Fax:81-836-22-2415)
COMMENT On Aug 31, 2000 this sequence version replaced gi:6172286
gi:6172288 gi:6172310 gi:6172312 gi:6172314 gi:6172316 gi:6172318
gi:6172320 gi:6172322 gi:6172324 gi:6635158 gi:6635160 gi:6635162
gi:6635164 gi:6635166 gi:6635168 gi:6635170 gi:6635172 gi:8947422
gi:8547426 gi:8547430 gi:8978373.
AB033780-AB033781, AB033792-AB033799: Submitted (25-Oct-1999)
AB033792-AB033799: Submitted (14-Feb-2000)
AB036071-AB036078: Submitted (18-Dec-2000).
FEATURES
Location/Qualifiers
1..299650
/organism="Chlamydomophila pneumoniae J138"
/strain="J138"
/db_xref="taxon:138677"
/ncbi_synonym="Chlamydia pneumoniae (strain J138)"
complement(1)..282
/gene="CpJ0001"
complement(1)..282
/gene="CpJ0001"
/codon_start=1
/transl_table=11
/product="CT001 hypothetical protein"
/protein_id="BAA98211.1"
/db_xref="GI:8978374"
/translation="MLGRIIRGLSLIVILCALNVGLIGITHNKLNIIAKLCGGVSTP
ATOITYIIIGIAGVICLLSFCPCSKKSHSHGSDSCSGGCHSHSDKN"
573..878
/gene="gatC"
573..878
/gene="gatC"
/codon_start=1
/transl_table=11
/product="Glu-tRNA Gln amidotransferase (C subunit)"
/protein_id="BAA98212.1"
/db_xref="GI:8978375"
/translation="MEQHLDRREILLAKASALQSEELIOEYQTSLSAVITSMKEA
LAIEDDADSCSESLFMHVNVNEDLREDSVTSDFNREFLRNVPESLGLVKVPAVK"
895..2373
/gene="gatA"
895..2373
/gene="gatA"
/codon_start=1
/transl_table=11
/product="Glu tRNA Gln amidotransferase (A subunit)"
/db_xref="GI:8978376"
/translation="MYRYSALAKAVTIGELTATGVTQHFHRIEABGGVQAFISL
CKEALQEALEIDKRSRGEPLGKLVGPKDNIHVTKLTCTCASRVLENYQPPFD
ATVVERIKEDGIIILKLNMFAMGSTTLYSAFHTHPNDLSRVPGSSGGSAAY
SARPCPVLGSDTGSIRQPAACGVGVGKPSYGAVSRYGLVAFASLLDQIGPLANTV
EDVALMDVFSGRDPKDATSRFFRDFSEMSKLTSEVPKVGVPRTFLEGLRDDIRENF
FSSALFEGEGTHLVDELIDLSHAVSIYIILASAEATNLARFCGVRYGYSPOAHT
ISQLYDLRSGEGFVEMRRIRLGNVLSAERQNYIKKATRAKRIYKAFRTAPEKC
EILAPQSSPAFEIGEILDPTVLYQDIYTVAMNLAYLPAIYAVSPGSFSGLEPLGLQ
IIGCQDQCVQGVGVSFQEAHQIKQLFSKRYAKSVVLGGQS"
2370..3836
/gene="gatB"
2370..3836
/gene="gatB"
/codon_start=1
/transl_table=11
/product="Glu tRNA Gln amidotransferase (B
subunit)"
/protein_id="BAA98214.1"

```


OM of: US-09-677-752-4 to: Issued Patents_NA:* out_format : pfs

Date: May 6, 2001 8:35 PM

About: Results were produced by the GenCore software, version 4.5.

Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+p2n.model
-Q/cgn2_1/USPTO.spool/US0967752/runat_04052001_132443_4939/app_query.fasta_1.948
-DB=Issued Patents_NA -ORMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -XGAPOP=10.000 -YGAPEXT=0.500
-TRANS=human40.cdi -LIST=45 -START=1 -MATRIX=blosum62
-THR_MAX=100 -THR_MIN=0 -ALIGN=25 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -MINLEN=0 -MAXLEN=200000000
-USER=US0967752/cgn1_1_70 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1
```

Search information block:

```
Query: US-09-677-752-4
Query length: 878
Database: Issued Patents_NA:*
Database sequences: 302621
Database length: 87301344
Search time (sec): 68.530000
```

score_list:

Sequence	Strd Orig	Zscore	Escore	Len	Documentation
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-038-682-3	148.50	227.85	4.2e-05	4937	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-302-832-3	148.50	227.85	4.2e-05	4937	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-530-198-3	148.50	227.85	4.2e-05	4937	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-469-880-3	148.50	227.85	4.2e-05	4937	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-728-470-3	148.50	227.85	4.2e-05	4937	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-617-697-6	148.50	218.60	0.0001	9323	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-302-832-6	148.50	218.60	0.0001	9323	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-530-198-6	148.50	218.60	0.0001	9323	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-469-880-6	148.50	218.60	0.0001	9323	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-728-470-6	148.50	218.60	0.0001	9323	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-617-697-6	145.00	221.43	9.5e-05	4794	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-038-682-7	144.00	221.11	9.9e-05	4287	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-302-832-7	144.00	221.11	9.9e-05	4287	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-530-198-7	144.00	221.11	9.9e-05	4287	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-469-880-7	144.00	221.11	9.9e-05	4287	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-728-470-7	144.00	221.11	9.9e-05	4287	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-212-133A-7	141.00	207.11	0.0006	7493	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-474-503-5	141.00	207.11	0.0006	7493	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-670-707A-5	141.00	207.11	0.0006	7493	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-037-601-5	141.00	207.11	0.0006	7493	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-038-682-1	139.00	208.76	0.0005	5116	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-302-832-1	139.00	208.76	0.0005	5116	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-530-198-1	139.00	208.76	0.0005	5116	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-469-880-1	139.00	208.76	0.0005	5116	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-728-470-1	139.00	208.76	0.0005	5116	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-617-697-1	137.00	202.64	0.0011	5952	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-663-998-2	137.00	202.64	0.0011	5952	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-320-559-1	135.50	187.00	0.0079	14255	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-327-392-1	135.50	187.00	0.0079	14255	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-306-691B-55	135.50	187.00	0.0079	14255	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-545-860D-1	135.50	187.00	0.0079	14255	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-169-927-1	134.50	199.39	0.0016	5319	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-038-682-5	134.00	190.48	0.0050	9171	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-302-832-5	134.00	190.48	0.0050	9171	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-530-198-5	134.00	190.48	0.0050	9171	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-469-880-5	134.00	190.48	0.0050	9171	

```
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-728-470-5 + 134.00 190.48 0.0050 9171
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-617-697-5 + 134.00 190.48 0.0050 9171
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-687-956A-22 + 132.00 195.18 0.0028 5077
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-457-273B-41 + 132.00 184.82 0.0104 1034
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-556-419-13 + 132.00 184.82 0.0104 10348
```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-038-682-3

```
seq_documentation_block:
: Sequence 3, Application US/08038682
: Patent No. 5549897
: GENERAL INFORMATION:
: APPLICANT: BARENKAMP, STEPHEN J
: APPLICANT: ST. GEME III, JOSEPH W
: TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
: TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Mattare, Ltd
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
```

```
COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/038,682
: FILING DATE: 16-MAR-1993
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: BERKSTRESSER, JERRY W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-293
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4937 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-038-682-3
```

```
alignment_scores:
: Quality: 148.50 Length: 811
: Ratio: 0.408 Gaps: 35
: Percent Similarity: 44.883 Percent Identity: 17.879
```

US-08-038-682-3

alignment_block:

US-09-677-752-4 x US-08-038-682-3

Align seg 1/1 to: US-08-038-682-3 from: 1 to: 4937

```
73 AsnSerTyCyTtrPheValSerLysLeuHisIleThrAspProLysgl 89
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
778 AACGACGAAGTCCTTTTTCATCAACCAAAATGATATCAAGTAGGTAAAGA 827
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89 uAlaLeuPheLysGluLysGly.....AspLeuSerI 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
828 CGCAATTATTACACTAATGCTTTACGCTTCTACGTAGACATTCTA 877
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100 LeGlnAsnPheArgPheLeuSerPheThrAspCysSerLysGluSer 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
878 ACGAAACATCAAGCGCGTATTTTCACCTTCGACGACCAACCAAGATAAA 927
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```

2488 AACAAGGAGCGAAAGTAATTTCAATTAACCAACCAACGAGACATGAA 2537
584 nThrTyrSerAspMetGlnAlaValGlnSerMetIleAsnThrIleAlaH 601
2538 CACA...AGCAAACTTTACCAATTCGGTTTGTAGCCAAATATCACAGCCA 2584
601 isGlyGlyAlaTyrLeuPheGlyThrTrpGlySerAlaValSerAsnLeu 617
2585 CTGGTGGG.....GGCTGTGTTTTTTTGTATATA 2613
618 PheTyrAlaHisSerSerGlyLysProIleAsp...AsnTrpHisH 633
2614 TATGCCAACCATCTGCCAGAGGGCTGAGTTAAATGAGTGAATTA 2663
633 sArgSerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAspH 650
2664 TATCTCTAACGGCGCTAATTTTACCTTAATTCCTATGTTCCGCGCGATG 2713
650 isSerPheCysLeuAlaAlaGlyGlnLeuLeuGlyLysSerSerAspSer 666
2714 ACGCTTTTAAATC.....AACAAAGACTTA 2739
667 PheIleThrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaG 683
2740 ACCATAATGCACCAATTC..... 2760
683 nLeuAlaThrSerLeuMetLysIleSerAlaGlnAlaCystYrAsnGluS 700
2761 .....AATTCA 2767
700 erIleHisGluLeuLysThrLys...TyrArgSerPheSerLysGluGly 715
2768 GCCTCAGACAGCAAGAGATGATTTTATGACGGGTACGACGCAATGCC 2817
716 PheGlySerTrpHisSerValAlaValSerGly 726
2818 ATCAATTCACTTACACATATCAATTCCTGGG 2850
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:us-08-302-832-3
seq_documentation_block:
; Sequence 3, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W

```

```

; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-302-832-3

alignment_scores:
Quality: 148.50 Length: 811
Ratio: 0.408 Gaps: 35
Percent Similarity: 44.883 Percent Identity: 17.879

alignment_block:
US-09-677-752-4 x US-08-302-832-3
Align seg 1/1 to: US-08-302-832-3 from: 1 to: 4937

73 AsnSerTyrCysTrpPheValSerLysLeuHisIleThrAspProLysGI 89
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
778 AACGGACAAGTCTTTTATCAACCCAAATGGTATCACAAATAGTAAAGA 827

89 uAlaLeuPheLysGluLysGly.....AspLeuSerI 100
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
828 CGCAATTATTAAACACTAATGGCTTTACGGCTTCTACGCTAGACATTCTA 877

100 leGlnAsnPheArgPheLeuSerPheThrAspCysSerSerLysGluSer 116
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
878 ACCAAACATCAAGCCCGCTAATTTCCCTTCGAGCAACCAAGATAAA 927

117 SerProSerIleIleHisGlnLysAsnGlyLysLeuSerLeuArgAsnS 133
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
928 GCGCTCGCTGAATTT...GTGAATCAGCGTTTAAATCTACTCGGTAAAGA 974

133 nGlySerMetSerPheCysArgAsnHisAlaGluGlySerGlyGlyAlaI 150
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
975 CGGCAGCTGTAATCTTATT.....GGTGGCAAG 1003

150 leSerAlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1004 TGA..... 1008

167 GluGluAsnSerSerLysGlyAsnGlyGlyAlaIleGln.....AI 180
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1009 AACGAGGGTGTGATTAGCGTAATGGTGGCAGCATTTCTTTACTCGCAGG 1058

180 aGlnThrPheSerLeuSerArgAsnValSerPro.....I 192
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1059 GCAAAAAATCACCATCAGCGATATAATAAACCAACCACTTACTTACAGCA 1108

192 leSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCysCys 208
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1109 TTCCCGCGCTGAAATGAAGCG...GTCAATCTGGCGGATATTTTGGCC 1155

209 SerAsnLeuIleCysSerGlyAsnValAsnProLeuPhePheThrGlyAs 225
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1156 AAA.....GGCGGTACATTAAATGTC.....CG 1178

225 nSerAlaThrAsnGlyGlyAlaIleCysCysIleSerAspLeuAsnThrS 242
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1179 TGCTGCCACT..... 1188

242 erGluLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAla 258
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1189 .....ATTGAAACCAAGCTAAACTTTCTGCT 1215

```

2818 ATCAATTACCTACACATATCCATTCTGGGC 2850

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-530-198-3

seq_documentation_block:

: Sequence 3, Application US/08530198

: Patent No. 5869085

: GENERAL INFORMATION:

: APPLICANT: BARENKAMP, STEPHEN J

: APPLICANT: ST. GEME III, JOSEPH W

: TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS

: TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS

: NUMBER OF SEQUENCES: 8

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Shoemaker and Mattare, Ltd

: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

: CITY: Arlington

: STATE: Virginia

: COUNTRY: U.S.A.

: ZIP: 22202-0286

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: PatentIn Release #1.0, Version #1.25

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/530,198

: FILING DATE: 13-DEC-1995

: CLASSIFICATION: 424

: ATTORNEY/AGENT INFORMATION:

: NAME: BERKSTRESSER, JERRY W

: REGISTRATION NUMBER: 22,651

: REFERENCE/DOCKET NUMBER: JWB-1186

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (703) 415-0810

: TELEFAX: (703) 415-0813

: INFORMATION FOR SEQ ID NO: 3:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 4937 base pairs

: TYPE: nucleic acid

: STRANDEDNESS: single

: TOPOLOGY: linear

: MOLECULE TYPE: DNA (genomic)

: US-08-530-198-3

alignment_scores:

Quality:	148.50	Length:	811
Ratio:	0.408	Gaps:	35
Percent Similarity:	44.883	Percent Identity:	17.879

alignment_block:

US-09-677-752-4 x US-08-530-198-3

Align seg 1/1 to: US-08-530-198-3 from: 1 to: 4937

```
73 AsnSerTyrCysTyrPheValSerLysLeuHisIleThrAspProLysG1 89
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
778 AACGGACAAAGTCTTTTATCAACCCCAATGGTATCAACAATAGGTAAAGA 827
89 uAlaLeuPheLysGluLysGly.....AspLeuSerI 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
828 CCAATTATACACTAATGGCTTTACGGCTTCTACGCTAGACATTCTTA 877
100 leGlnAsnPheArgPheLeuSerPheThrAspCysSerLysGluSer 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
878 ACGAAACATCAAGCGGGGTATTTACCTTCGAGCAACCAAGATAAA 927
117 SerProSerIleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAs 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
928 GCGCTCGTGAATTT...GTGAATCACGGTTTAAATTACTGTCGGTAAAGA 974
```

```
133 nGlySerMetSerPheCysArgAsnHisAlaGluGlySerGlyGlyAlaI 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
975 CGCAGTGTAAATCTTATT.....GGTGGCAAG 1003
150 leSerAlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1004 TGAAA..... 1008
167 GluGluAsnSerSerLysGlyAsnGlyGlyAlaIleGln.....Al 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1009 AACGAGGTGCTGATTACCGTAAATGGTGGCAGCATTTCTTTACTCGCAG 1058
180 aGlnThrPheSerLeuSerArgAsnValSerPro.....I 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1059 GCAAAATACCATCATCGCATTAATAAACCCCAACCATTTACTTACAGCA 1108
192 leSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCysCys 208
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1109 TTGCGCGCGCTGAAATGAAGCG...GTCAATCTGGCGGATATTTTGCC 1155
209 SerAsnLeuIleCysSerGlyAsnValAsnProLeuPhePheThrGlyAs 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1156 AAA.....GCGGTACATTAAATGTC.....CG 1178
225 nSerAlaThrAsnGlyGlyAlaIleCysCysIleSerAspLeuAsnThrS 242
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1179 TGCTGCCACT..... 1188
242 erGluLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAla 258
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1189 .....ATTCGAAACCAAGGTAAACTTTCTGCT 1215
259 SerAsnSerAlaLysGluLysGlyGlyAlaIleTyr.....AlaLysHI 273
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1216 GATTCTGTAGCAAGATAAAGCGCAATATTGTTCTTTCCGCCAAGA 1265
273 sMetValLeuArgTyrAsnGlyProValSerPheIleAsnAsnSerAlaL 290
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1266 GGGTCAACGGCAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTA 1315
290 ys.....IleGlyGlyAlaIleAlaIleGlnSerGly 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1316 AAGCGGCAAGCTGATGATTACAGCGGATAAGTCACATTAAACACAGT 1365
301 GlySerLeuSerIleLeuAlaGlyGluGlySerValLeuPheGlnAsnAs 317
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1366 GCAGTTATCGACCTTTCAGGTAAAGAGGGGAGAACTTACCTTGGCGG 1415
317 nSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeuG 334
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1416 TCACGAGCGC.....GCGAAGGTAAACACGGCATTTCAATTAG 1453
334 LuLysAspAlaIleLeuSerSerLeuGluAlaArgAsn..... 346
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1454 CAAGAAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGGCCAAA 1503
347 .....GlyAspIleLeuPhePheAspPr 354
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1504 CAAAAAGCGGACGCGCTATTGTGTGGGCGGATATTGCTTAATTCAC... 1551
354 oLleValGlnGluSerSerLysGluSerProLeuProSerSerLeuG 371
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1552 .....GCCAATATTA 1561
371 InAlaSerValThrSerProThrProAlaThrAlaSerProLeuValIle 387
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1562 ACGCTCAAGTAGTGTGATATCGCTAAACCGGTGGT.....TTTGTG 1605
388 GlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerG1 404
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1606 GAGACATCGGGCATTATTTATCCATTGACAGCAATGCAATTGTTAAAC 1655
404 uGluGlu.....LysThrProAspAsnLeuThrSerGlnLeuGlnGlnP 419
```

REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 4937 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-469-880-3

alignment_scores:

Quality: 148.50 Length: 811

Ratio: 0.408 Gaps: 35

Percent Similarity: 44.883 Percent Identity: 17.879

alignment_block:

US-09-677-752-4 x US-08-469-880-3 ..

Align seg 1/1 to: US-08-469-880-3 from: 1 to: 4937

```

73 AsnSerTyrCysTrpPheValSerLysLeuHisIleThrAspProLysGI 89
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
778 AACGGACAAGCTCTTTTAAATCAACCCAAATGGTATCACAAATAGTAAGA 827
89 uAlaLeuPheLysGluLysGly.....AspLeuSerI 100
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
828 CGCAATTATTAAACACTAATGGCTTTACGGCTTCTACGGTAGACATTCTA 877
100 LeuGlnAsnPheArgPheLeuSerPheThrAspCysSerLysGluSer 116
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
878 ACGAACATCAGCAGCGGTAATTTACCTTCGAGCAACCAAGATATAA 927
117 SerProSerIleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAs 133
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
928 CGCGCTCGCTGAAATT...GTGAATCAGCGTTTAAATTACTGTCGTAAGA 974
133 nGlySerMetSerPheCysArgAsnHisAlaGluGlySerGlyGlyAlaI 150
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
975 CGCAGTGTAATCTTATT.....GCTGGCAAG 1003
150 leSerAlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1004 TGAAG..... 1008
167 GluGluAsnSerSerLysGlyAsnGlyGlyAlaIleGln.....A 180
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1009 AACGAGGGTGTATTAGCGTAAATGGTGGCAGCATTCCTTACTCGCAGG 1058
180 aGlnThrPheSerLeuSerArgAsnValSerPro.....I 192
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1059 GCAAAATAATCACCACGAGCATATAATAACCCACCATTTACTACAGCA 1108
192 leSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCysCys 208
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1109 TTGCGCGGCTGAAATGAGCG...GTCNAATCTGGCGATATTTTGGCC 1155
209 SerAsnLeuIleCysSerGlyAsnValAsnProLeuPhePheThrGlyAs 225
||| : : : : : : : : : : : : : : : : : : : : : : : : :
1156 AAA.....GGCGGTAACATTAAATGTC.....CG 1178
225 nSerAlaThrAsnGlyAlaIleCysCysIleSerAspLeuAsnThrS 242
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1179 TCGTGCACCT..... 1188
242 erGluLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAla 258
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1189 .....ATTGGAACCAAGGTAACCTTCTGCT 1215
259 SerAsnSerAlaLysGluLysGlyGlyAlaIleTyr.....AlaLysHi 273

```

```

1216 GATTCTGTAAGCAAGATAAAAGCGCAATATTGTTCTTTCCGCCCAAGA 1265
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
273 sMetValLeuArgTyrAsnGlyProValSerPheIleAsnAsnSerAlaL 290
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1266 GGGTGAACGCGCAAAATGGCGGTGTAAATTCCTCCGCTCAAAATCAGCAAGCTA 1315
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
290 ys.....IleGlyGlyAlaIleAlaIleGlnSerGly 300
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1316 AAGGCGCAAGCTGATGATTACAGGCGCATAAAGTCACATTAAACACAGGT 1365
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
301 GlySerLeuSerIleLeuAlaGlyGluGlySerValLeuPheGlnAsnAs 317
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1366 GCAGTTATCCACCTTTTCAGGTAAAGAGGGGAGAAACTTTACCTTGGCGG 1415
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
317 nSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeuG 334
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1416 TGACGAGCGC.....GCCGAAGGTAAACCGCATTCATTCAATTAG 1453
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
334 LuLysAspAlaIleLeuSerSerLeuGluAlaArgAsn..... 346
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1454 CAAAGAAACCTCTTTAGAAAAAGGCTCAACCATCATGTATCAGGCAAA 1503
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
347 .....GlyAspIleLeuPhePheAspPr 354
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1504 GAAAAAGGCGGACGCGCTATTGTGTGGGCGGATATTGCGTTAATTGAC... 1551
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
354 olleValGlnGluSerSerSerLysGluSerProLeuProSerSerLeuG 371
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1552 .....GGCAATATTA 1561
371 lnAlaSerValThrSerProThrProAlaThrAlaSerProLeuValIle 387
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1562 ACCTCAAGGTATGTGTGATATCGCTAAACCCGGTGT.....TTTGTG 1605
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
388 GlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerGI 404
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1606 GACACATCGGGGCAATTATTTATCAATTGACAGCAATGCAATTTGTTAAAC 1655
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
404 uGluGlu.....LysThrProAspAsnLeuThrSerGlnLeuGlnP 419
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1656 AAAAGAGTGGTTCCTACACCGCTGATGATGTAACAATTTGAAGCCCAAGACC 1705
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
419 rolleGluLeuLysSerGlyArgLeuValLeuLysAspArg..... 432
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1706 CCCTTCGCANTATACCGGT.....ATAATGATGAATCCCNACA 1746
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
433 .....AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLe 447
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1747 GGCACCGGTCAAGCAAGCGACCTTAAATAAATAGCGAACTCAAAACAAC 1796
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
447 uLeuIle..... 449
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1797 GCTAACCAATAACAATAATTTCATAAATATCTGAAAAACGCGCTGGACAATGA 1846
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
450 .MetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAla 465
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1847 ATATACGGCATCAAGAAAACCTTACCGTTAATAGCTCAATCAACATCGGA 1896
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
466 ThrLeuSer...IleProLeuHisSer..... 473
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1897 AGCAACTCCCACTTAATTTCTCCATAGTAAAGGTGAGCGTGGCGGAGCGGT 1946
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
473 ..... 473
1947 TCAGATTGATGGAGATATTACTTTCTAAAGCGGAAATTTAACCATTTATT 1996
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
474 .....LeuAspThrGluLysSerValThrIleHisAlaProAsn 486
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1997 CTGCGCGGATGGTGTATGTTTCATAAAAATATTACGCTTGATCAGGGTTT 2046
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
487 LeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspGluAsnPheTy 503
||||| : : : : : : : : : : : : : : : : : : : : : : : : :

```



```
1189 .....ATTCGAACCAAGGTAAACTTCTGCT 1215
259 SerAsnSerAlaLysGluLysGlyGlyAlaIleTyr.....AlaLysHi 273
1216 GATTCTGTAGCAAGAGTAAAGCGGCAATATTGTTCTTTCCGCCAANAGA 1265
273 sMetValLeuArgTyrAsnGlyProValSerPheIleAsnAsnSerAlaL 290
1266 GGGTGAAGCGGAATTCGGCGGTGTAAATTCGGCTCAAAATCAGCAAGCTA 1315
290 ys.....IleClyAlaIleAlaIleGlnSerGly 300
1316 AAGCGGCAAGCTGATGATTACAGGCGCATTAAGTCACATTAAAAACAGGT 1365
301 GlySerLeuSerIleLeuAlaGlyGlySerValLeuPheGlnAsnAs 317
1366 GCAGTTATCGACCTTCAGGTAAAGAGGGGAGNAACCTTACCTTGGCGG 1415
317 nSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeuG 334
1416 TGACGAGCGC.....GGCGAAGGTAAACAGCGCATTCAATTAG 1453
334 LuLysAspAlaIleLeuSerSerLeuGluAlaArgAsn..... 346
1454 CAAGAAACCTCTTTAGAAAAAGGCTCAACCATCAATGTATCAGGCAAA 1503
347 .....GlyAspIleLeuPhePheAspPr 354
1504 GANAAGCGGCGCGCTATTGTGTGGCGGATATTCGGTTAATTGAC.. 1551
354 oIleValGlnGluSerSerLysGluSerProLeuProSerSerLeuG 371
1552 .....GGCAATATTA 1561
371 InAlaSerValThrSerProThrProAlaThrAlaSerProLeuValle 387
1562 ACGCTCAAGGTAGTGGTGATATCGCTAAAAACCGGTGGT.....TTGTG 1605
388 GlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerG 404
1606 GAGACATCGGGCAATATTATCCATTGACAGCAATGCCAATGTTAAAC 1655
404 uGluGlu.....LysThrProAspAsnLeuThrSerGlnLeuGlnP 419
1656 AAAAGAGTGGTGTAGACCTGTATGATGATTAACAATTCAGCCGCAAGACC 1705
419 rolleGluLeuLysSerGlyArgLeuValLeuLysAspArg..... 432
1706 CCCTTCGCAATAATACCGGT.....ATAATGATGAATTCGCCAACA 1746
433 .....AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLe 447
1747 GGCACCGGTGAGCAGCAGCCCTTAAAAAANTAGCGAATCAAAACAC 1796
447 uLeuIle..... 449
1797 GCTAACCAATACAACATTTTCAAAATTTATCTGAAAAACGCGCTGGACAATGA 1846
450 ..MetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAla 465
1847 ATATAACGGCATCAAGAAACTTACCGCTTAATAGCTCAATCAACATCGGA 1896
466 ThrLeuSer...IleProLeuHisSer..... 473
1897 AGCAACTCCNCTTAATCTCCATGTAAGGTGAGCGTGGCGGAGCGGT 1946
473 ..... 473
1947 TCAGATTGATGAGATATTACTTCTAAAGCGGCAAAATTTAACCAATTATT 1996
474 .....LeuAspThrGluLysSerValThrIleHisAlaProAsn 486
.....
```



```
2043 GTGGAGACATCGGGGCATTATTTATCCATTACAGCAATGCAATTGTAA 2092
403 rGluGluGlu.....LysThrProAspAsnLeuThrSerGlnLeuGlnG 418
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2093 ACAAAAGAGTGGTGTAGACCTGATGATGATACCAATTGAAGCCGAAG 2142
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
418 lnProfileGluLeuLysSerGlyArgLeuValLeuLysAspArg..... 432
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2143 ACCCCCTTCGCAATAATACCGGT.....ATAAATGATGAATTCCTCA 2183
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
433 .....AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAl 446
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2184 ACAGGCACCGGTGAAGCAAGCGACCTTAAATAAATAAGCAATCAAAAC 2233
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
446 aLeuLeuLe..... 449
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2234 AACCTAACCAATACAACTATTTCAAAATATCTGAAAAAGCCCTGGACAA 2283
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
450 .....MetGluAlaGlyThrSerLeuLysThrSerAspLeuLysLeu 464
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2284 TGAATAATACGGCATCAAGAAACTTACCGTTAATAGCTCAATCAACATC 2333
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
465 AlaThrLeuSer.....IleProLeuHisSer..... 473
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2334 GGAAGCAACTCCCACTTAATTTCTCATAGTAAAGGTCAGCGTGGCGAGG 2383
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
473 ..... 473
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2384 CGTTCAGATTGATGGAGATATTCTTAAGCGCGGAATTTAACCAATTT 2433
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
474 .....LeuAspThrGlyLysSerValThrIleHisAlaPro 485
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2434 ATTCGCGCGATGGTGTGATGTTTCATAAAATATTACGCTTGATCAGGT 2483
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
486 AsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspGluAsnPh 502
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2484 TTTTAAATATTACCGCCCGCTTCGCTAGCTTTGAAGGTGGAATAACAA 2533
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
502 eTyrGlu.....AsnValGluLeuLeuSerLys..... 511
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2534 AGCAGCGCGCGCAATGCTAATATTGCGCCCGGCACTGTACCA 2583
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
512 .....GluGlnAsnAsnIleProLeu..... 518
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2584 TTACAGGAGAGGAAAGATTTCAGGCTAACACGATCTTTAAACGGA 2633
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
519 .....LeuThrLeuSerLysGluGlnSerHisLeuHisLeuPr 531
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2634 ACGGTAAGGTCTGAATATCATTTTCATCAGTGAATAATTTA..... 2675
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
531 oAspGlyAsnLeuSerSerHisPheGlyTyrGlnGlyAspTrpThrPheS 548
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2676 .ACCCACAATCTTAGTGGCACATTAACATATCTGGGAATATAACAATTA 2724
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
548 er.....TrpLysAspSerAspGluGly 555
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2725 ACCAACTACGAGAAAGACACCTCGTATTGGCAACACCGCCATGATTCG 2774
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
556 HisSerLeuLe.....AlaAsnTrpThrPr 564
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2775 CACTGGACGTCAGTGTCTTAATCTAGACACAGCGCAAAATTTACCTT 2824
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
564 oLysAsnTyrValProHisProGluArgGlnSerThr..... 576
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2825 TATTAAATACATTTCAAGCAATAGCAAGGCTTAACCAACACAGTATAGAA 2874
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
576 ..... 576
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2875 GCTCTGCGGGGTGAATTTTAACGGCGTAATGCAACATGTCATTCAAT 2924
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
577 .....LeuValAlaAsnThrLeuTr 583
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2925 CTCAAGCAAGGAGGCAAGTTAATTTCAAAATTAACCAACCAAGCAACAT 2974
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-302-832-6

```
seq_documentation_block:
: Sequence 6, Application US/08302832
: Patent No. 5603938
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Mattare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/302,832
: FILING DATE: 16-SEP-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US pct/us93/02166
: FILING DATE: 16-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstreser, Jerry W
: REGISTRATION NUMBER: 22,651
```

```
seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq;US-08-530-198-6

seq_documentation_block:
; Sequence 6, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530.198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: JWB-1186
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9323 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-530-198-6

alignment_scores:
Quality: 148.50 Length: 812
Ratio: 0.408 Gaps: 36
Percent similarity: 44.828 Percent identity: 18.227

alignment_block:
US-09-677-752-4 x US-08-530-198-6 ..
Align seg 1/1 to: US-08-530-198-6 from: 1 to: 9323

73 AsnSerTyrCysTrpPheValSerLysLeuHisIleThrAspProLysG1 89
||||| : : : : : ||||| : : : : :
1218 ACGGCACAGCTCTTTTAATCAACCCAAATGGTATCACATAAGTAAAGA 1267
||||| : : : : : ||||| : : : : :
89 uAlaLeuPheLysGluLysGly.....AspLeuSerI 100
||||| : : : : : ||||| : : : : :
1268 CGCAATTATTAACTAATAATGGCTTTACGGCTTCACGTAGACATTTCTA 1317
||||| : : : : : ||||| : : : : :
100 leGlnAsnPheArgPheLeuSerPheThrAspCysSerSerLysGluSer 116
||||| : : : : : ||||| : : : : :
1318 ACGAACAACATCAAGCGGCCGTAAATTTACCCTTCGAGCAAAACCAAGATAA 1367
||||| : : : : : ||||| : : : : :
117 SerProSerIleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAs 133
||||| : : : : : ||||| : : : : :
1368 GCCTCGCTCAAAATT...GTGAATCACGGTTTTAATTACTGTCGTAAGA 1414
||||| : : : : : ||||| : : : : :
133 nGlySerMetSerPheCysArgAsnHisAlaGluGlySerGlyAlaI 150
||||| : : : : : ||||| : : : : :
```

600 laHisGlyGlyAlaTyrLeuPheGlyThrTrpGlySerAlaValSerAsp 616
|| |||||
3022 CCACTGGTGG.....GGCTGTGTTTTTTTGAT 3050

617 LeuPheTyrAlaHisAspSerSerGlyLysProIleAsp...AsnTrpHi 632
: : : : : ||| : : : : :
3051 ATATATGCCAACCACTTCGACAGGGCTGAGCTTAAATAGATGATAAT 3100
: : : : : ||| : : : : :
632 sHisArgSerLeuGlyTyrLeuPheGlyThrHisSerLeuAspA 649
: : : : : ||| : : : : :
3101 TAATATCTCTAACGGCGCTAAATTTTACCTTAAATCCCATGTTCCGGCGG 3150

649 sHisSerPheCysLeuAlaAlaGlyClnLeuLeuGlyLysSerSerAsp 665
|| : : : : :
3151 ATAGCGCTTTTAAATC.....AACAAAGAC 3176
: : : : : ||| : : : : :
666 SerPheIleThrSerThrGluThrThrSerTyrIleAlaThrValGlnAl 682
|| : : : : : ||| : : : : :
3177 TTAACCATAAATCCACCAATTC..... 3200

682 aGlnLeuAlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrAsnG 699
|||
3201AATT 3204

699 luSerIleHisGluLeuLysThrLys...TyrArgSerPheSerLysGlu 714
||| : : : : : ||| : : : : :
3205 TCAGCTCTACAGACAGAAAGATGATTTTATGACGGGTACGACGCAAT 3254

715 GlyPheCysSerThrHisSerValAlaValSerGly 726
: : : : : ||| : : : : :
3255 GCCATCAATTCAACCTACACATATCATCTCTGGC 3290

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:us-08-469-880-6

seq_documentation_block:
; Sequence 6, Application US/08469880
; Patent No. 5876733
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Shoemaker and Matlare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,880
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-516 MIS.v9

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9323 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-469-880-6

alignment_scores:
Quality: 148.50 Length: 812
Ratio: 0.408 Gaps: 36
Percent Similarity: 44.828 Percent Identity: 18.227

alignment_block:
US-09-677-752-4 x US-08-469-880-6

Align seg 1/1 to: US-08-469-880-6 from: 1 to: 9323

73 AsnSerTyrCysTrpPheValSerLysLeuHisIleThrAspProIysGI 89
||||| : : : : : |||||
1218 AACGGACAACTCTTTTAAATCAACCCAAATGGTATCACAAATAGTAAAGA 1267

89 uAlaLeuPheLysGluLysGly.....AspLeuSerI 100
||||| : : : : : |||||
1268 CGCAATTATTAACTAATGGCTTACGGCTTCTACGCTAGACATTCTTA 1317

100 leGlnAsnPheArgPheLeuSerPheThrAspCysSerSerLysGluSer 116
||||| : : : : : |||||
1318 ACGAAACATCAAGCGCGTAATTTTACCTTCGAGCAACCAAGATAAAA 1367

117 SerProSerIleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAs 133
: : : : : ||| : : : : : |||||
1368 CGCGCTCGCTCAAAATT...GTGAATCAGCGTTTAAATTACTGTCGTAAGA 1414

133 nGlySerMetSerPheCysArgAsnHisAlaGluGlySerGlyAlaI 150
||||| : : : : : |||||
1415 CGCAGCTGCTAAATCTTATT.....GGTCGCAAG 1443

150 leSerAlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
: : : : :
1444 TGAAT..... 1448

167 GluGluAsnSerSerLysGlyAsnGlyGlyAlaIleGln.....Al 180
: : : : : ||| : : : : : |||||
1449 AACGAGGTGTGATTACCGTAAATGGTGGCAGCATTTCTTACTCGCAGG 1498

180 aGlnThrPheSerLeuSerArgAsnValSerPro.....I 192
||| : : : : : ||| : : : : :
1499 GCAAAAAATCACCATCAGCATATATAAACCACCACTTACTTACAGCA 1548

192 leSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCysCys 208
: : : : : ||| : : : : : |||||
1549 TTGCCGCGCTGAAATGAAGCG...GTCAATCTGGCGCATATTTTGGCC 1595

209 SerAsnLeuIleCysSerGlyAsnValAsnProLeuPheThrGlyAs 225
: : : : : ||| : : : : : |||||
1596 AAA.....GCCGTACATTAAATGTC.....CG 1618

225 nSerAlaThrAsnGlyGlyAlaIleCysCysIleSerAspLeuAsnThr 242
: : : : : ||| : : : : : |||||
1619 TGCTGCCACT..... 1628

242 erGluLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAla 258
|||||
1629ATCGAAACCAAGGTAACTTTCTGCT 1655

259 SerAsnSerAlaLysGluLysGlyGlyAlaIleTyr.....AlaLysHi 273
: : : : : ||| : : : : : |||||


```

; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9323 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-617-697-6

alignment_scores:
  Quality: 148.50      Length: 812
  Ratio: 0.408        Gaps: 36
Percent Similarity: 44.828 Percent Identity: 18.227

alignment_block:
US-09-677-752-4 x US-08-617-697-6 ..
Align seg 1/1 to: US-08-617-697-6 from: 1 to: 9323

73 AsnSerTyrCysTrpPheValSerLysLeuHisIleThrAspProLysG1 89
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1218 AACGCACAAAGTCCTTTTAAATCAACCCAAATGGTATCACATAGTAAGA 1267
89 uAlaLeuPheLysGluLysGly.....AspLeuSerI 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1268 CGCAATTATTACACTATGCGCTTACGGCTTCTACGCTAGACATTCTA 1317
100 leGlnAsnPheArgPheLeuSerPheThrAspCysSerLysGluSer 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1318 ACGAAACATCAAGCGCGTAATTTACCTTCGAGCAACCAAGATAAA 1367
117 SerProSerIleIleHisGlnLysAsnGlyGlnLeuSerLysArgAsnAs 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1368 CGCGTCGCTGCAATTT...GTCAATCACGGTTTAAATTACTGCGTAAAGA 1414
133 nGlySerMetSerPheCysArgAsnHisAlaGluLysSerGlyAlaI 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1415 CGGCAGGTGTAATCTTATT.....GGTGGCAAG 1443
150 leSerAlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1444 TGAAG..... 1448
167 GluGluAsnSerSerLysGlyAsnGlyGlyAlaIleGln.....Al 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1449 AACGAGGCGTGATTACGCTAAATGGTGGCAGCATTTCTTACTCGCAGG 1498
180 aGlnThrPheSerLeuSerArgAsnValSerPro.....I 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1499 GCAAAATACCATCAGCATATTAATAAACCCCAACCATTTACTACAGCA 1548
192 leSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCysCys 208
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1549 TTGCGCGCGCTGAAATCAACGCG...GTCAATCTGGCGGATATTTTGCC 1595
209 SerAsnLeuIleCysSerGlyAsnValAsnProLeuPheThrGlyAs 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1596 AAA.....GGCGGTAAACATTAAATGTC.....CG 1618
225 nSerAlaThrAsnGlyGlyAlaIleCysCysIleSerAspLeuAsnThrS 242
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1619 TGCTGCCACT..... 1628
242 erGluLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAla 258
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

1629 .....ATTGAAACCAAGGTAAACTTTTCTGCT 1655
259 SerAsnSerAlaLysGluLysGlyGlyAlaIleTyr.....AlaLysHi 273
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1656 GATTCTGTAGCAAGATAAGCGCAATATTGTCTTTCCGCCAAAGA 1705
273 smetValLeuArgTyrAsnGlyProValSerPheIleAsnAsnSerAlaL 290
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1706 GGTGAACGCGAAATTCGGGTGTAAATTTCCGCTCAAAATCAGCAAGCTA 1755
290 ysIleGlyGlyAlaIleAlaIleGlnSer..... 299
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1756 AA...GGCGCAAGCTGATGATAAAGTCCGATAAAGTCACATTAAAAACA 1802
300 GlyGlySerLeuSerIleLeuAlaGlyGluGlySerValLeuPheGlnAs 316
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1803 GGTGCAGTTATCGACCTTTTCAGGTAAAGAGGGGAGAAACTTACCTTGG 1852
316 nAsnSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrL 333
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1853 CGGTGACGAGCGC.....GGCAAGGTAAAAACGCGATTCAAT 1890
333 euGluLysAspAlaIleLeuSerSerLeuGluAlaArgAsn..... 346
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1891 TAGCAAGAAAAACCTCTTTACAAAAAGGCTCAACCATCAATGTATCAGGC 1940
347 .....GlyAspIleLeuPhePheAs 353
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1941 AAGAAAAAGCGGACGCGCTATTGTGTGGGGCGATATTGCGTTAAATTGA 1990
353 pProIleValGlnGluSerSerLysGluSerProLeuProSerSerL 370
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1991 C.....GGCAATA 1998
370 euGlnAlaSerValThrSerProThrProAlaThrAlaSerProLeuVal 386
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1999 TTAACGCTCAAGGTAGTGGTGATATCGCTAAACCGGTGGT.....TTT 2042
387 IleGlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSe 403
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2043 GTGAGACATCGGGCATTTATTTATCCATTGACAGCAATGCAATTGTATA 2092
403 rGluGluGlu.....LysThrProAspAsnLeuThrSerGlnLeuGlnG 418
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2093 AACAAAGAGTGGTGTAGACCTGATGATACAAATTGAAGCCGAAG 2142
418 lnProIleGluLeuLysSerGlyArgLeuValLeuLysAspArg..... 432
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2143 ACCCCCTTCGCAATAATACCGGT.....ATAAATGATGAATTCCCA 2183
433 .....AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAl 446
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2184 ACAGGCACCGGTGAAGCAAGCGACCCCTAAAAAAATAGCGAACTCAAAAC 2233
446 aLeuLeuIle..... 449
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2234 AACGCTACCAATACAACTATTTCAAATTATCTGAAAAACGCTGGACAA 2283
450 .....MetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeu 464
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2284 TGAATATAACGGCATCAAGAAAACTTACCGTTAATAGCTCAATCAACATC 2333
465 AlaThrLeuSer...IleProLeuHisSer..... 473
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2334 GGAAGCAACTCCCACTTAATTTCTCATAGTAAAGTCAGCGTGGCGAGG 2383
473 ..... 473
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2384 CGTTCAGATTGATGAGATATTACTTTCTAAAGCGGGAATTTAACCATTT 2433
474 .....LeuAspThrGluLysSerValThrIleHisAlaPro 485
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2434 ATTCTGGCGGATGGTTTCATGTTTCATAAAATATTACCGCTTCATCAGGT 2483

```

527 ACAGAAACATCAAGCGCGTAATTTCCACCTTGAGCAAAACCAAGGATAAA 576
 117 SerProSerIleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAs 133
 577 GCATCTGCTGAAATC...GTGAATCAGCGTTTAAATTACCGTTGGTAAAGA 623
 133 nGlySerMetSerPheCysArgAsnHisAlaGluGlySerGlyGlyAlaI 150
 624 CGGTAGCGTAAACCTTATT...GGTGCAAG 652
 150 LeSerAlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
 653 TGA... 657
 167 GluGluAsnSerSerLysGlyAsnGlyGlyAlaIleGln...Al 180
 658 AACGAGCGGTGATTAGCGTAATTCGGGTAGTATTCTTTTACTTTCGAGG 707
 180 aGlnThrPheSerLeuSerArgAsnValSerPro...I 192
 708 GCAAAATCACCATCAGCGATATAATAATCCAAACCATCACATCACAGCA 757
 192 LeSerPheAlaArgAsnAlaAspLeuAsnGlyGlyAlaIleCysCys 208
 758 TTGCTGCACCTGAAACGAGCG...ATCAATCTGGCGATATTTTGGC 804
 209 SerAsnLeuIleCysSerGlyAsnValAsnProLeuPheThrGlyAs 225
 805 AAA...GTGGTAAACATTAAT... 822
 225 nSerAlaThrAsnGlyGlyAlaIleCysCysIleSerAspLeuAsnThrS 242
 823 ...GTCCGCGCTGCCACTATTC 841
 242 eGluLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAla 258
 842 GCAATAAAGTAAACTTTCT...GCC 864
 259 SerAsnSerAlaLysGluLysGlyGlyAlaIleTyr...AlaLysHI 273
 865 GACTCTGTAGCAAGATAAAGTGGTAACATTTCTCTCTGCGCAAGA 914
 273 sMetValLeuArgTyrAsnGlyProValSerPheIleAsnAsnSerAlaL 290
 915 AGGTGAACCGGAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCCA 964
 290 ys...IleGlyGlyAlaIleAlaIleGlnSerGly 300
 965 AAGGTGGTAAGTTGATGATTACAGCGGTAATAAGTTACATTGAAACCGGT 1014
 301 GlySerLeuSerIleLeuAlaGlyGluGlySerValLeuPheGlnAsnAs 317
 1015 GCATGTATCGACCTTCGGGTAAGAGAGGGGAGAAACTTATCTTGGCGG 1064
 317 nSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeuG 334
 1065 TGACCGCGT...GCGGAGGTAAACCGGCATTCAATTAG 1102
 334 LuLysAspAlaIleLeuSerSerLeuGluAlaArgAsn... 346
 1103 CAAGAAACACCTTTAGAAAAGGCTCAACAATTAATGTGTCAGGTAAA 1152
 347 ...GlyAspIleLeuPhePheAspPr 354
 1153 GAAAAAGTGGCGCGTATTTGATGGCGGATATTCGGTTAATTTGACGG 1202
 354 oIleValGlnGluSerSerLysGluSerProLeuProSerSerLeuG 371
 1203 CAATATT...AATGCC 1216
 371 InAlaSerValThrSerProThrProAlaThrAlaSerProLeuValIle 387
 1217 AAGGTAAAGATATCGCTAAACACTGGTGT...TTTGTG 1251

388 GlnThrSerAla...AsnArgSerValIlePheSe 398
 1352 CAGACGTGGGGCATTTATTCATTCATGATGAACCGCAATTTGTTAAAC 1301
 398 rSerGluArgLeuSerGluGluGlyThrProAspAsnLeuThrSerG 415
 1302 AAAAGAATGGCTACTAGAC...CCAGAGAATGTGACTATTG 1339
 415 InLeuGlnGlnProIleGluLeuLysSerGlyArgLeuValLeuLysAsp 431
 1340 AAGCTCTCTCCGTTCTCGCGTCGAGCTGGT...GCCGAT 1377
 432 ArgAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeuLe 448
 1378 AGGAATTCCTCCACTCGGAGAGGTGATAAA...GTGAC 1412
 448 uIleMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuA 465
 1413 CCTAAAAAATAAACACCTCTTTGACACACTAACCAATACACCATTT 1462
 465 laThrLeuSerIleProLeuHis...SerLeuAspThrGluLysSer 479
 1463 CAATCTTCTGAAAAGTCCACCGTGTGAACATAACGGCAAGGAGAAA 1512
 480 ValThrIleHisAlaProAsnLeuSerIleGlnLys...IlePh 493
 1513 CTTACCTCTTAATAGC...TCTATCAGTATAGAAAGGCTCCCACTTAAT 1559
 493 eLeuSerAsnSerGlyAspGluAsnPheTyrGluAsnValGluLeu... 508
 1560 TCTCCACAGTCAGGTGAGGCGGT...CAAGGTGTTTCAGATTGATA 1603
 509 ...LeuSerLysGluGlnAsnAsnIleProLeuLeuThrLeuSerLys 523
 1604 AAGATATTACTTCTGAAGCGGAAATTTAACCAATTTATCTGCGGATGG 1653
 524 GluGlnSerHisLeuHisLeuProAspGly...AsnLeuSe 536
 1654 GTTCATGTTTCAATAAAATATTACCTTTGGTAGCGGCTTTTAAACATCAC 1703
 536 rSerHisPheGlyTyrGlnGlyAspTyrThrPheSerTrpLysAspSerA 553
 1704 AACTAAA...GAAGGAGATATCGCTTCGAAGACAGTCTGGAC 1744
 553 spGluGlyHisSerLeuIleAlaAsnTrpThrProLysAsnTyrValPro 569
 1745 GGAACAACCTAACCATTAACGCCAAGGAGGCATCAGCTCAGGTAATAGT 1794
 570 HisProGluArgGlnSerThrLeuValAlaAsnThrLeuTrpAsn... 584
 1795 AACGGCTTTTATGATTTAAACAACGCTCTCTAAACAGCCTTGCGGAAAGCT 1844
 585 ThrTyrSerAspMet...GlnAlaValGlnSerMetI 596
 1845 GAGCTTTTACTGACAGCAGAGGACAGGTAGAGAACTAAGGGTATA 1894
 596 leAsnThrIleAlaHisGlyGlyAlaTyrLeuPheGlyThrTrpGlySer 612
 1895 TCTCAACAATAATTTGACGGAACGTTAAACATTTCCGGAAT... 1935
 613 AlaValSerAsnLeuPheTyrAlaHisAspSerSerGlyLysProIleAs 629
 1936 ...GTAGATATCTCAATGAAGACCCCAAGT 1964
 629 AsnTrpHisHisArgSerLeuGlyTyrLeuPhe...GlyIleSerThrH 645
 1965 CAGCTGGTTTTACAGACAAAGACCGCACTACTGGAACGTAAACACTT 2014
 645 isSerLeuAspAspHisSerPheCysLeuAlaAlaGlyGlnLeuLeuGly 661
 2015 TAAATGTT... 2022


```
371 nAlaSerValThrSerProThrProAlaThrAlaSerProLeuValIleG 388
2230 AGTAATAATTACCAAT.....C 2246
388 lnThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerGlu 404
2247 AGATAACAATCAAGCAGCGCTTCAAGGGGGCGTCTGTGATCAAGTGAG 2296
405 GluLulysThrProAspAsnLeuThrSerGlnLeuGlnProIleG1 421
2297 GCAGAAATGCT.....ACCTAACTATTCAACACCAA.....GA 2331
421 uLeuLys...SerGlyArgLeuValLeuLys.....AspArgAlaValL 435
2332 GTTAAATTTGGCAGGACCACTAATATTTCAGCGTTTAAATAAGCAGAAA 2381
435 euSerAlaProSerLeuSerGlnAspProGlnAlaLeuLeuIleMetGlu 451
2382 TTACAGCTAAATAATGGCAGTGAT.....TTAACT 2410
452 AlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAlaThr..... 466
2411 ATTGGCAATGCTAGCGGTGGTATGCTGCTAAATAAGTGACTTTTGA 2460
467 .....LeuSerIleProLeuHisSerLeuAspThrG 477
2461 CAAGGTTAAGATTCAAAATCTCGACTGAGGTCACAATGTACACTAA 2510
477 luysservAlThrIleHisAlaProAsnLeuSerIleGlnLysIlePhe 493
2511 ATACCGAAGTG.....AAAACGCTCTAATGCTAGT..... 2539
494 LeuSerAsnSerGlyAspGluAsnPheTyrGluAsnValGluLeuLeuSe 510
2540 ...AGCAATGCTGGTAAATGATAAC...AGACCGGTTTAAACCATTTCCG 2583
510 rLysGlu.....GlnAsnAsnIleProLeuLeuThrLeuSerLysG 524
2584 ARAAGATGTAAACGTTAAACAATACGTT.....ACCTCCACCA 2621
524 luGlnSerHisLeuHisLeuProAspGlyAsnLeuSerSerHisPheGly 540
2622 AGACAATAATATCTCTGCCGACGAGAAATGTAAACAACAAAGAGGC 2671
541 TyrGln...GlyAspTrpThrPheSerTrpLysAspSerAspGluGlyH 556
2672 ACAACTATCAATGCCAACCCAGGCGAGTGAAGTAACTGCTCAAAATGG 2721
556 sSerLeuIleAlaAsnTrpThrProLysAsnTyrValProHisProGluA 573
2722 TACAAATTAAGGCACAACTTACCTCGCAAAATGTAAACAGTACAGCAACAG 2771
573 rgGlnSerThrLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMet 589
2772 AAAATCTTGTACCACAGAGAATGCTGCTCATTAATGCAACCGCGGACA 2821
590 GlnAlaValGlnSerMetIleAsnThrIleAlaHisGlyAlaTyrLe 606
2822 GTAACCATTAAGTACAAAACAGGGGATATTAAAGGTGGA.....AT 2862
606 uPheGlyThrTrpGly.....SerAlaValSerAsnLeuPheT 619
2863 TGAATCAACTTCGGTGAATGTAATATTACAGCGGCAATACACTTA 2912
619 yrAlaHisAspSerSerGlyLysProIleAspAsnTrpHisHisArgSer 635
2913 AGCTAAGTAAATATACACTGGTCAAGATGTA..... 2941
636 LeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspHisSerPh 652
2942 .....ACAGTAACAGCGGATCCAGGAGCCTT 2967
```

```
652 eCysLeuAlaAlaGlyGlnLeuLeuGlyLysSerSerAspSerPheIleT 669
2968 GACAACTACAGCAGGCTCAACCATTTAGTCGCAACAACAGCAATGCAATA 3017
669 hrSerThrGluThrSerTyrIleAlaThrValGlnAlaGlnLeu... 684
3018 TTACACCAAAACAGGTGATATCAACGGTAAAGTTGAATCCAGCTCCGGC 3067
685 AlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrAsnGluSerI 701
3068 TCTGTAACACTTGTGCACTGGAGCAACTCTTGCTGTA..... 3106
701 eHisGluLeuLysThrLysTyrArgSerPheSerLysGluGlyPheGly 718
3107 .....GGTAATATTT 3116
718 erTrpHisSerValAlaVal.....SerGlyGluValCysAlaSer 731
3117 CAGTAACACTGTTACTATTACTGCGGATAGCGGTAATTAACCTCCACA 3166
732 IleProIleValSerAsnGlySerGlyLeuPheSerSerPheSerIlePh 748
3167 GTAGGTCTTACAATTAATGGGACTAATAGTAAACCCTCAAGCCAATC 3216
748 eSerLysLeuGlnGly...PheSerGlyThrGlnAspGlyPheGluGlu 764
3217 AGCGGATATTGAAGGTACAAATTTCTGGTAATACAGTAATGTTACAGCAA 3266
764 erSerGlyGluIle 768
3267 GCACCTGCTGATTGA 3280
```

seq_name: /cqn2.6/ptodata/2/ina/5B_COMB.seq:US-08-530-198-7

seq_documentation_block:
: Sequence 7, Application US/08530198
: Patent No. 5869065
: GENERAL INFORMATION:
: APPLICANT: BARENKAMP, STEPHEN J
: APPLICANT: ST. GEME III, JOSEPH W
: TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
: TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Maltare, Ltd
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/530,198
: FILING DATE: 13-DEC-1995
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: BERKSTRESSER, JERRY W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: JWB-1186
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4287 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear


```

STATE: Georgia
COUNTRY: US
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/212,133A
FILING DATE: March 11, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: EMU/76677
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-572-6508
TELEFAX: 404-572-6555
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 7493 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1..407
OTHER INFORMATION: /rpt_type="terminal"
OTHER INFORMATION: /note="5'UTR"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7471..7476
OTHER INFORMATION: /function="PolyA_signal"
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 7368..7493
OTHER INFORMATION: /rpt_type="terminal"
OTHER INFORMATION: /note="3'UTR"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 408..7367
OTHER INFORMATION: /product="Coagulation Factor VIII"
PUBLICATION INFORMATION:
AUTHORS: Elder, F.
AUTHORS: Laxich, D.
AUTHORS: Gitschier, J.
TITLE: Sequence of the Murine Factor VIII cDNA.
Patent No. 5663060
JOURNAL: Genomics
VOLUME: 16
PAGES: 374-379
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 7: FROM 1 TO 7476
US-08-212-133A-7

```

```

alignment_scores:
Quality: 141.00 Length: 753
Ratio: 0.387 Gaps: 39
Percent Similarity: 48.340 Percent Identity: 21.248

```

```

alignment_block:
US-09-677-752-4 x US-08-212-133A-7

```

```

Align seg 1/1 to: US-08-212-133A-7 from: 1 to: 7493
46 PheGluAspCysThr.....MetGluSerLeuPheProAl 57
||||| |||
2730 TTCAAAGATTCCACAATCCAAAAAATGATATGGAGAGATTGAGCCTCA 2779
57 aLeu.....CysAlaHisAlaSerGlnAspA 66
||||| |||
2780 GTTTGAAGAGATAGCAGAGATGCTTAAAGTACAGAGTGTCTCAGTTAGTG 2829
66 spProLeuTyrValLeuGlyAsnSerTyrCysTrpPheValSerLysLeu 82
||||| |||
2830 ACATGTTGATGCTCTTGGGACAGAGTCATCT...ACTCCACATGGCTTA 2876
83 HisIleThrAspProLysGluAlaLeuPheLys..... 93
||||| |||
2877 TTTTATCAGATGCCCAAGAGCCATCTATCAGGCTATTTCATGATGATCA 2926
94 .....GluLysGlyAspLeuSerIleGlnAsp 103
||||| |||
2927 TTCACCAATGCAATAGACAGCAATGAAGCCCATCTAAAGTAGACCAAC 2976
103 heArgPheLeuSerPheThrAspCysSerSerLysGluSerSerProSer 119
||||| |||
2977 TCAGG.....CCAGAATCCCATCAGTCAGAGAA 3005
120 IleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMe 136
||||| |||
3006 ATAGTATTACTCTCTCAGCCGGCTCCAGTTAAGATCCATATAAAAGTTT 3055
136 tSerPheCysArgAsnHisAlaGluGlySerGlyAlaIleSerAlaA 153
||||| |||
3056 G.....GAGACAACTATAGAAGTAAAGTGAAGAAACTTGGTTGC 3096
153 spAlaPheSerLeuGlnHisAsnTyrLeuPheThr..... 164
||||| |||
3097 AAGTTTCTAGTTGCCAAGTAATCTAATGACTACACAACAAATTTCTCTCAGAC 3146
165 .....AlaPheGluGlu..... 168
||||| |||
3147 AATTGAAAGCACTTTTGAAGACAGACAGATTCTTCAGGATTTCCAGATAT 3196
169 .....AsnSerSerLysGlyAsnGlyGlyAlaIleGlnAlaGlnT 182
||||| |||
3197 GCCAGTTCACTCTAGTAGTAATTAAGTACTACTGTCATTTGCTAAGAAAG 3246
182 hrPheSerLeuSerArgAsnValSerProIleSerPheAlaArgAsnArg 198
||||| |||
3247 CATATTCCTTGTGGTCTCTCATCTACCTTTTAAACGCGAGTGAAGAAAT 3296
199 AlaAspLeuAsnGlyGlyAlaIleCysSerSerAsnLeuIleCysSerG 215
||||| |||
3297 AGTGAT.....TCCAACATATTGGATTCA... 3320
215 yAsnValAsnProLeuPhePheThrGlyAsnSerAlaThrAsnGlyGlyA 232
||||| |||
3321 .....ACTTTAATGTATAGTCAAGAAAGTTTACCAAGAGATATA 3360
232 laIleCysCysIleSerAspLeuAsnThrSerGluLys.....GlySer 246
||||| |||
3361 TATTATCAATAGAGAAATAGATTACTCAGAGAGAGAGAGAGGTTTCATGGA 3410
247 LeuSerLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaL 263
||||| |||
3411 AITGCTTTATTGACCAAGATANTACTTTTCAAGACAATGTCCTT 3460
263 sGluLysGlyGlyAlaIleTyr.....AlaLysHisMetValLeuA 277
||||| |||
3461 AATGAACAACAACAACATATATATTCATTCAACAACATAATGAAAACTAC 3510
277 rgTyrAsnGlyProValSerPheIleAsnAsnSerAlaLysIleGlyGly 293
||||| |||
3511 AACTGAGAGGCCCAACATCAATTTGAGAATAGTACAAACAGACTTTGCAAGAT 3560

```



```

/ OTHER INFORMATION: /note= "5' UTR"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 7471..7476
/ OTHER INFORMATION: /function= "polyA signal"
/ FEATURE:
/ NAME/KEY: repeat_unit
/ LOCATION: 7368..7493
/ OTHER INFORMATION: /rpt_type= "terminal"
/ OTHER INFORMATION: /note= "3' UTR"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 408..7367
/ OTHER INFORMATION: /product= "coagulation factor VIII"
/ PUBLICATION INFORMATION:
/ AUTHORS: Elder, F.
/ AUTHORS: Lakich, D.
/ AUTHORS: Gitschler, J.
/ TITLE: Sequence of the murine Factor VIII cDNA
/ Patent No. 5859204
/ JOURNAL: Genomics
/ VOLUME: 16
/ PAGES: 374-379
/ DATE: 1993
/ US-08-670-707A-5

```

```

alignment_scores:
  Quality: 141.00      Length: 753
  Ratio: 0.387        Gaps: 39
  Percent Similarity: 48.340  Percent Identity: 21.248

alignment_block:
US-09-677-752-4 x US-08-670-707A-5

Align seg 1/1 to: US-08-670-707A-5 from: 1 to: 7493

46 PheGluAspCysThr.....MetGluSerLeuPheProAl 57
|||||:||||| ||| |||||:|||||: |||
2730 TTCGAAGATTCACAAATCCAAAAAATGATGGAGAAGATTGACCTCA 2779

57 aLeu.....CysAlaHisAlaSerGlnAspA 66
|||:|||||: ||| |||||: |||: |||
2780 GTTTCAGAGATAGCAGAGATGCTTAAAGTACAGAGTGTCTCAGTTAGTG 2829

66 sProLeuTyrValLeuGlyAsnSerTyrCysTyrPheValSerLysLeu 82
|| ||| |||:|||||: |||: |||
2830 ACATGTTGATGCTCTTGGGACAGAGTCATCCT...ACTCCACATGGCTTA 2876

83 HisIleThrAspProLysGluAlaLeuPheLys..... 93
|||||:|||||: |||: |||: |||: |||
2877 TTTTATCAGATGGCCAAAGACCCATCATGAGGCTATTATCATGATGATCA 2926

94 .....GluLysGlyAspLeuSerIleGlnAsnP 103
|||||:|||||: |||: |||: |||: |||
2927 TTCACCAATGCAATAGACAGCAATGAAGGCCCATCTAAAGTGACCCCAAC 2976

103 heArgPheLeuSerPheThrAspCysSerLysGlySerProSer 119
|||||: |||: |||: |||: |||: |||
2977 TCAGG.....CCAGATCCCATCATCAGATGAGAAA 3005

120 IleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMe 136
|||||: |||: |||: |||: |||: |||: |||
3006 ATAGTATTCTTCTCCTCAGCCCGGCTCCAGTTAAGATCCAAATAAAGTTT 3055

136 tSerPheCysArgAsnHisAlaGluGlySerGlyAlaIleSerAlaA 153
|||||: |||: |||: |||: |||: |||
3056 G.....GAGCAACTATAGAAAGTAAAGTGGGAAGAACTTGTGTTCC 3096

153 sPalaPheSerLeuGlnHisAsnTyrLeuPheThr..... 164
|||||: |||||: |||: |||: |||
3097 AAGTTTCTAGTTGCCAAGTATCTACTGACTACAACTTCTGTGTCAGAC 3146

```

```

165 .....AlaPheGluGlu..... 168
|||||: |||: |||: |||: |||
3147 AATTGAAAGCAACTTTTGAAGAGACACATCTCTCAGCATTTCCAGATAT 3196

169 .....AsnSerSerLysGlyAsnGlyGlyAlaIleGlnAlaGlnT 182
|||||: |||: |||: |||: |||
3197 GCCAGTTCACCTCTAGTAGTAATTAAGTACTACTGCTTTGGTAAAGAG 3246

182 hrPheSerLeuSerArgAsnValSerProIleSerPheAlaArgAsnArg 198
|||||: |||: |||: |||: |||
3247 CATATTCCCTTGTGGGTCTCATGTACCTTTAAACCGCAGGTGAAGAAAT 3296

199 AlaAspLeuAsnGlyGlyAlaIleCysSerAsnLeuIleCysSerG1 215
|||||: |||: |||: |||: |||
3297 AGTGAT.....TCCAACATATTGGATTCA... 3320

215 yAsnValAsnProLeuPhePheThrGlyAsnSerAlaThrAsnGlyGlyA 232
|||||: |||: |||: |||: |||
3321 .....ACTTTAATGTATAGTCAAGAAAGTTTACCAAGAGATAATA 3360

232 laIleCysCysIleSerAspLeuAsnThrSerGluLys.....GlySer 246
|||||: |||: |||: |||: |||
3361 TATTATCAATAGAGAATCATAGATTACTCAGAGAGAAGAGGTTTCATGGA 3410

247 LeuSerLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaL 263
|||||: |||: |||: |||: |||
3411 ATTGCTTTATTGACCAGAGATAATATCTTTTCAAGAGACAATGCTCCTT 3460

263 sGluLysGlyGlyAlaIleTyr.....AlaLysHisMetValLeuA 277
|||||: |||: |||: |||: |||
3461 AATGAAACAAACAAACATATAATCATTCAACAACTAATGAAAAACTAC 3510

277 rgTyrAsnGlyProValSerPheIleAsnAsnSerAlaLysIleGlyGly 293
|||||: |||: |||: |||: |||
3511 ACATGAGAGCCCAACATCAATTGAGAATAGTACACACAGACTTGCAAGAT 3560

294 AlaIle.....AlaIleGlnSerGlyGlySerLeuSerI1 305
|||||: |||: |||: |||: |||
3561 CCCATATTAAAGGTCAATAGTCAGATTCAA.....GAAGTAACAGC 3601

305 eleuAlaGlyGluGlySerValLeuPheGlnAsnAsnSerGlnArgThrS 322
|||||: |||: |||: |||: |||
3602 TTGATTCTATGATGACACATTTTAGGCAAAAATTTCTACA..... 3641

322 erAspGlnGlyLeuValArgAsnAlaIleTyrLeuGluLysAspAlaIle 338
|||||: |||: |||: |||: |||
3642 .....TATTGAGACTAAACCATATG 3662

339 LeuSer.....SerLeuGluAlaArgAsnGlyAspIleLeuPhe..... 351
|||||: |||: |||: |||: |||
3663 CTAATAGACTACTCTCAACAAAAATAAGACATATTTCATAGAAAAAGA 3712

352 .PheAspProIleValGlnGluSerSerLysGlySerProLeuProS 368
|||||: |||: |||: |||: |||
3713 TGAAGTTCCTATTCCACAAGATGAAGAGAATAACAATCATGCCATTTTCCA 3762

368 erSerLeuGlnAlaSerValThrSerProThrProAlaThrAlaSerPro 384
|||||: |||: |||: |||: |||
3763 AGATGTTGTTCTGTGCA.....GAATCTTCAAT 3791

385 LeuValIleGlnThrSerAlaAsnArgSerValIlePheSerSerGluAR 401
|||||: |||: |||: |||: |||
3792 TCGTTTAAAAACACCAATCGAAATAATTCCTTC..... 3824

401 gluSerGluGluGluLysThrProAspAsnLeuThrSerGlnLeuGlnG 418
|||||: |||: |||: |||: |||
3825 .AACTCTGAGCAAGAACATAGTCCAAAGCAATTAGTATATTTAATG.... 3869

418 InProIleGluLeuLysSerGlyArgLeuValLeuLysAspArgAlaVal 434
|||||: |||: |||: |||: |||
3870 .....TTTAAAAAATATGTAAAAAATCAAAAGTTTC 3899

435 LeuSerAlaPro.....SerLeuSerGlnAspProGlnAlaLeuLe 448

```

```

;
;   AUTHORS:  Elder, F.
;   AUTHORS:  Lakich, D.
;   AUTHORS:  Gitschier, J.
;   TITLE:    Sequence of the murine Factor VIII CDNA
;   Patent No. 6180371
;   JOURNAL:  Genomics
;   VOLUME:   16
;   PAGES:    374-379
;   DATE:     1993
;   US-09-037-601-5

alignment_scores:
    Quality: 141.00      Length: 753
    Ratio:   0.387       Gaps: 39
    Percent Similarity: 48.340    Percent Identity: 21.248

alignment_block:
US-09-677-752-4 x US-09-037-601-5 ..
Align seg 1/1 to: US-09-037-601-5 from: 1 to: 7493

46  PheGluAspCysThr.....MetGluSerLeuPheProAl 57
|||||  |||  |||||  |||  |||
2730 TTCAAAGATTCCCAAAATCCCAAAATGATATGGAGAAGATTGAGCCTCA 2779

57 aleu.....CysAlaHisAlaSerGlnAspa 66

2780 GTTTGAAGAGATACAGAGATGCTTAAAGTACAGAGTCTCTCAGTTAGT 2829

66 spProLeuTyrValLeuGlyAsnSerTyrCysTrpPheValSerLysLeu 82
|||  |||  |||||  |||  |||
2830 ACATGTCGTCCTCTGGGACAGAGTCATCCT...ACTCCACATGGCTTA 2876

83 HisIleThrAspProLysGluAlaLeuPheLys..... 93
|||||  |||  |||||  |||  |||
2877 TTTTATCAGATGGCCAGAGAGCCATCTATGAGGCTATTTCATGATGATCA 2926

94 .....GluLysGlyAspLeuSerIleGlnAsnP 103

2927 TTCACCAATGCATAGACACATGATGAAGGCCATCTAAAGTGACCCCAAC 2976

103 heArgPheLeuSerPheThrAspCysSerSerLysGluSerSerProSer 119
|||||  |||  |||||  |||  |||
2977 TCAGG.....CCAGAATCCCATCAGAGTGAGAAA 3005

120 IleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMe 136
|||||  |||  |||||  |||  |||
3006 ATAGTATTACTCTCAGCCCGGCTCCAGTTAAGATCCCAATAAAGTTT 3055

136 lserPheCysArgAsnHisAlaGluGlySerGlyClyAlaIleSerAlaA 153
|||||  |||  |||||  |||  |||
3056 G.....GAGACAATATAGAGTAAAGTGAAGAAACTTGGTTGC 3096

153 spAlaPheSerLeuGlnHisAsnTyrLeuPheThr..... 164
|||||  |||  |||||  |||  |||
3097 AAGTTTCTAGTTGGCAAGTATCTAATGACTACACAACTTCTGTCAGAC 3146

165 .....AlaPheGluGlu..... 168
|||||  |||  |||||  |||  |||
3147 AATTTGAAAGCAACTTTTGAAGAGACAGATCTTCCAGGATTTCAGATAT 3196

169 .....AsnSerSerLysGlyAsnGlyClyAlaIleGlnAlaGlnT 182
|||||  |||  |||||  |||  |||
3197 GCCAGTTCACCTAGTAGTAATTAAGTACTACTGCTATTTGGTGAAGAAAG 3246

182 hrPheSerLeuSerArgAsnValSerProIleSerPheAlaArgAsnArg 198
|||||  |||  |||||  |||  |||
3247 CATATTCCCTTGTGGGTCTCATGTACCTTTAAACCGGAGTGGAAGAAAT 3296

199 AlaAspLeuAsnGlyClyAlaIleCysCysSerAsnLeuIleCysSerG1 215
|||||  |||  |||||  |||  |||
3297 ACTGAT.....TCCAACATATTGGATTCA... 3320

```

```

215 yAsnValAsnProLeuPheThrGlyAsnSerAlaThrAsnGlyGlyA 232
|||||  |||  |||||  |||  |||
3321 .....ACTTTATGTATAGTCAAGAAAGTTTACCAAGAGATATA 3360

232 lalleCysCysIleSerAspLeuAsnThrSerGluLys.....GlySer 246
|||  |||  |||||  |||  |||
3361 TATTATCAATAGAGAATCATAGATTACTACAGAGAGAGAGGTTTTCATGGA 3410

247 LeuSerLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLy 263
|||||  |||  |||||  |||  |||
3411 ATTGCTTTATTGACCAAGAGATACTATTATTCAAAGACAAATGCTCCTT 3460

263 sGluLysGlyClyAlaIleTyr.....AlaLysHisMetValLeuA 277
|||  |||  |||||  |||  |||
3461 AATGAAAACAACAACAACATATATCATCTCAACAACATAAGTAAAAACTAC 3510

277 rgTyrAsnGlyProValSerPheIleAsnAsnSerAlaLysIleGlyCly 293
|||  |||  |||||  |||  |||
3511 ACACGTGAGAGCCCAACATCAATTTCAGAATAGTACAACAGAGACTTCAAGAT 3560

294 AlaIle.....AlaIleGlnSerGlySerLeuSerIle 305
|||||  |||  |||||  |||  |||
3561 GCCATATTAAAGGTCAATAGTGAAGATTCAA.....GAAGTAACAGC 3601

305 eLeuAlaGlyGluGlySerValLeuPheGlnAsnAsnSerGlnArgThrS 322
|||  |||  |||||  |||  |||
3602 TTTGATTCTGATGACACTTTTAGCCAAAATTTCTACA..... 3641

322 erAspGlnGlyLeuValArgAsnAlaIleTyrLeuGluLysAspAlaIle 338
|||||  |||  |||||  |||  |||
3642 .....TATTTGAGACTAAACCATATG 3662

339 LeuSer.....SerLeuGluAlaArgAsnGlyAspIleLeuPhe..... 351
|||||  |||  |||||  |||  |||
3663 CTAATAGAACTAGCTCAACAAAAAATAAGACATATTTTCATAGAAAAAGA 3712

352 .PheAspProIleValGlnGluSerSerSerLysGluSerProLeuProS 368
|||||  |||  |||||  |||  |||
3713 TGAGATTCCTATTCCACAGATGAAGAGAAATACATCATGCCATTTTCCA 3762

368 erSerLeuGlnAlaSerValThrSerProThrProAlaThrAlaSerPro 384
|||  |||  |||||  |||  |||
3763 AGATGTTCTTCTTGTC.....GAATCTTCAAAAT 3791

385 LeuValIleGlnThrSerAlaAsnArgSerValIlePheSerSerGluAr 401
|||||  |||  |||||  |||  |||
3792 TGGTTTAAAAAGACCAATGGAATAATTTCTCTG..... 3824

401 gLeuSerGluGluGlyThrProAspAsnLeuThrSerGlnLeuGlnG 418
|||||  |||  |||||  |||  |||
3825 .AATCTGAGCAGAGAACATAGTCCAAACCAATTAGTATTATTAATG.... 3869

418 lnProIleGluLeuLysSerGlyArgLeuValLeuLysAspArgAlaVal 434
|||  |||  |||||  |||  |||
3870 .....TTTAAAAAATATGTAAAAAATCAAAAGTTTC 3899

435 LeuSerAlaPro.....SerLeuSerGlnAspProGlnAlaLeuLe 448
|||||  |||  |||||  |||  |||
3900 TTGTCAGAGAAAAATAAAGTCACAGTAGAACAGGATGATTTTCAACAGAA 3949

448 uIleMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuA 465
|||  |||  |||||  |||  |||
3950 CATAGGACTTAAAGACATGGCTTTTCCACATAATATGACATATTTCTTA 3999

465 laThrLeuSerIleProLeuHisSer.....LeuAspThrGluLys 478
|||||  |||  |||||  |||  |||
4000 CCACCTTGTCT....AACGTACATGAAATGGTAGCCACAATCAAGAAAAA 4046

479 SerVal.....ThrIleHisAlaProAsnLeuSerIleGlnLysIlePh 493
|||||  |||  |||||  |||  |||
4047 AATATTCAGAGAGATAGAGAAGCAAGCACTAATTGAAGAGAAAGTAGT 4096

```



```

312 ValLeuPheGlnAsnSerGlnArgThrSerAspGlnGlyLeuValAr 328
1059 .....CAAATAATCACCATCAGCGGATATAATAA 1087
328 gAsnAlaIleTyrLeu.....GluYAspAlaIle...L 339
1088 CCCAACCACTTACTTACAGATTGCGCGCTGAAATGAACGGGTCAATC 1137
339 euSerSerLeuGluAlaArgAsnGlyAspIleLeuPhePheAspProile 355
1138 TGGCGGATATTTCGCCAAGCGGTACATT..... 1169
356 ValGlnGluSerSerSerLysGluSerProLeuProSerSerLeuGlnAl 372
1169 ..... 1169
372 aserValThrSerProThrProAlaThrAlaSerProLeuValIleGlnT 389
1169 ..... 1169
389 hrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerGluGlu 405
1170 ..AATGTCGGTGGTCCACTATTGCAACCAAGGTAAACTTTCGCTCAT 1217
406 GluTyrThrProAspAsnLeuThrSerGlnLeuGlnProIleGluLe 422
1218 TCTGTAAGCAAGAT..... 1232
422 uLysSerGlyArgLeuValLeuLysAspArg..... 432
1233 .AAAAGCGGCATATTGTTCTTCCGCCCAAGAGCGGTGAACGGGAATPG 1281
433 ..AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeuLeu 448
1282 GCGGTGTATTTCCCGCTCAAAATCAGCAGCTTAAGCGGCAAGCTGATG 1331
449 IleMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAl 465
1332 ATTACAGCGGATAAAGTCACTATAAAACAGAGGTGCGATTATCGACCTTC 1381
465 a..... 465
1382 AGGTAAGAGGGCGGAGAACTTACCTTGGCGGTGACGCGCGCGGAAG 1431
466 ..ThrLeuSerIleProLeu...HisSerLeuAspThrGluLysSerVal 480
1432 GTAAAGGGCATTCANTTAGCAAGCAAAACCTCTTTAGAAAAGGCTCA 1481
481 ThrIleHisAlaPro..... 485
1482 ACCATCATGTATCAGCAAGAAAGGCGGCGGCTATTGTGTGGG 1531
486 .....AsnLeuSerIleGlnLysIle..... 492
1532 CGATATTGCGTTAATTACGCGCAATATTAAACGCTCAAGGTAGTGTGATA 1581
493 .....PheLeuSerAsnSerGlyAspGluAsnPheTyr 503
1582 TCCTAAACCGGTGTTTGTGGAGCGCTGGGGCATGATTATTATCATC 1631
504 GluAsnValGluLeuLeuSerLysGluGlnAsnAsnIleProLeuLeuTh 520
1632 AAGACAATCAATTGTCAGCGCAAGAG..... 1661
520 rLeuSerLysGluGlnSerHisLeuHisLeuProAspGlyAsnLeuSerS 537
1662 .....TGGTTGTATGACCCCGAT...AATGTATCTA 1689
537 erHisPheGlyTyrGlnGlyAspTrpThrPheSerTrpLysAspSerAsp 553
1690 TTAATGCAGAAACAGCAGGACGCAATACTTCAGAACGACGATGATAC 1739
554 GluGlyHisSerLeuIleAlaAsnTrpThrProLysAsnTyrValProHI 570

```



```

4338 TACTACTCAACCTATTAACGGCGCTTTGG.....GACCAATTCA 4395
      ::::: ||:::||||| ||| :::::
590 nalavalinsermet.1leasnthrllealhisgllyalatyler 606
      :::||||:||||| :::::||||| :::::
4336 CACTGTCAACTAATAGCGCTTAACCATTAACGTTCACACAGTATAC 4445
      :::||||:||||| :::::||||| :::::
607 pheglythrtprglyseralavalserasleuphetryalahispsse 623
      :::
4446 ATT..... 4448
623 rserolytprolleasprntprhisarbserleuglytylerup 640
      ||| ||||| ||| :::::||||| :::::
4449 .....AACAAGATAGAC.....CACGCAAGAAATGAAAAACCTC 4484
640 heglyleaserthrhisserleuasprhisserphecysleualala 656
      |||::: |||||:::||||| ||| :::::
4445 .....ATTACTAATCTTCTCTATCATGATTTCTGTGTAAACACAC 4529
657 glycineleuglylserserasprserphelthserthrcluph 673
      ||| :::|||||:::||||| ||| :::::||||| :::::
4530 ACTCGAGCA.....ATAGTCTGTACTACACATTTGTAATAACATCAGC 4573
673 rthsertryllealathvalcinalacineleualathrsermet 690
      : ||| ::::: :::::
4574 ATTTCACCAATAGATCTCCAAAGACGTCATTCCAAAATAATTTCTC 4623
690 ysileselacinalacystyrasnaglserllehisleulelysthr 706
      ::::: |||:::|||||:::|||||
4624 ATGTTCAACGA.....TCATCTCAATTTATGACTTTAAGACA 4661
707 lys 707
|||
4662 AAA 4664

seq_name: /cgn2.6/ptodata/2/ina/5A_COMB.seq:US-08-038-682-1
seq_documentation_block:
  Sequence 1, Application US/08038682
  Patient No 5548897
  GENERAL INFORMATION:
  APPLICANT: BARENKAMP, STEPHEN J
  APPLICANT: ST. GENE III, JOSEPH W
  TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
  TITLE OF INVENTION: OF NON-TYPEABLE HAEMORRHOIDS
  NUMBER OF SEQUENCES: 8
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: Shoemaker and Matlare, Ltd
  STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
  CITY: Arlington
  STATE: Virginia
  COUNTRY: U.S.A.
  ZIP: 22202-0286
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: Patent Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/038.682
  FILING DATE: 16-MAR-1993
  CLASSIFICATION: 424
  ATTORNEY/AGENT INFORMATION:
  NAME: BERSSTRESSER, JERRY W
  REGISTRATION NUMBER: 22,651
  REFERENCE/DOCKET NUMBER: 1038-293
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: (703) 415-0810
  TELEFAX: (703) 415-0813
  INFORMATION FOR SEQ ID NO: 1:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 5116 base pairs

```

```

;      TYPE: nucleic acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
;      MOLECULE TYPE: DNA (genomic)
US-08-038-682-1

```

```
alignment_scores:
  Quality: 139.00      Length: 819
  Ratio: 0.364        Gaps: 40
Percent Similarity: 46.642  Percent Identity: 19.902
```

```
alignment_block:
US-09-677-752-4 x US-08-038-682-1 .
```

Align seg 1/1 to: US-08-038-682-1 from: 1 to: 5116

94 GLiuyScLySjAapLeuSerIleGln...AsnPhetArgPheLeuSerPheTh 109
470 GAAATCGCGTCACTTACGGTTAAACGCCATTCCGCTATGTTACTATTCCT 519
109 rAspCys.SerSerLySjGluSer..... 117
520 TAGCGTACATCTATTCACCAATTCGTTTAAAGCAGCGCTTACACAGA 569
118 ProSerIleIleHisGlnLySAsnGlyLeuSerLeuArgAsnAsnGly 134
570 ATGCGATGTAGTACAC.....GGCACACGCCACTATGCAAGTAGATGG 610
134 ySerMetSerPheCysArgAsnHisAlaGluGlySerGlyAlaIle 151
611 TAATTAACCCATTATTCGCCACAGCTGTTGAC.....GATATCATTA 651
151 eArgAlaSPAlaPheSerLeuGlnHisAsnTyPLeuPheThAlaPheGlu 167
652 ATTGGAACATTTTAACTTCGACCCAAATGAAATGGTCAGCTTTTACAA 701
168 GLiAsnSerSerLySjAsnGlyAlaIleGlnAlaGlnThrPheSe 184
702 GAAAC.....AACACATCCGCGCTATTAACCGCTGTATC 739
184 rLeuSerArgAsnValSerProIleSerPheAlaArgAsnArgAlaSP 201
740 T.....AACCAATCTCCCAATTAAAGGATTTTAGATT 774
201 euAsnGlyAlaIleLeuCysSerSerAsnLeuIleCysSerGlyAsnVal 217
775 CTAAAGCGACAGTCTTTTA.....ATC 797
218 AsnProLeuPhePheThGlyAsnSerAlaThrAsnGlyAlaIleCys 234
798 AACCA.....AATGCTATCAACAATAGG 820
234 sCysIleSerAspLeuAsnThrSerGlnGlySerLeuSerLeuAlaC 251
821 TAAAGACCACTATTATTAACACTAATAGCGCTTACGGCTTACGCTACA 870
251 yAsnGlnGluThrLeuPheAlaSerAsn.....SerAlaLys 263
871 TTTCTAAGAAACATCAAGCGCGCTATTCACCTGGAGCAACAA 920
264 GLiuyScLyGlyAlaIleTyAlaLysHis.....MetValLeuArgTy 278
921 GATAAAGCGCTCGCGGAATGTGCAATACGCTTAATTAATCTGCGTAA 970
278 rAsnGlyProValSerPheIleAsnAsnSerAlaLysIleGlyAlaI 295
971 AGACGCGAGTGTAAATCTTATTGCGGCAAGTGCAAAACACAGGCTGCA 1020
295 leaIleIleGlnSerGlySerLeuSerIleLeuAlaGlyGluGlySer 311
1021 TTAAAGGTA...AATGCTGCGACATTTCTTATCTTCGCGAGG..... 1058

```

493 eLeuSerAsn.....SerGlyaspGluAsnPhetYrGluAsnV 506
    ||| ..... |||
4097 TTGGCCCGGAGTGCAGACAGCACTGGCTTAAGATTCTTGAAAGACA 4146
    ||| ..... |||
506 aLgUleuLeuSerLysGluGlnAsn..... 514
    ||| ..... |||
4147 TAATGATAGTACGACTAGGCAAAATATAGTTATATAGAGTACAGTGA 4196
    ||| ..... |||
515 .....AsnIleProLeuThLeuSerLysGluGlnSerH1 527
    ||| ..... |||
4197 CCAGTACTTAAACATACATCATATAACATTACACAAATACAGTACA 4246
    ||| ..... |||
527 sLeuH1sLeuProAspG1YasnLeuSerSerH1sPhG1YTrGlnG1Y 544
    ||| ..... |||
4247 GATTCACATG.....GAGCATTTCTTTAAAGAAAGG. 4277
    ||| ..... |||
544 sPTPrThrPheserTrpLysAspseraspGluG1YH1sSerLeuI1A1a 560
    ||| ..... |||
4278 .....AGGACACAGAGAAACAAATTCAGAAAGCCCTTGTA 4310
    ||| ..... |||
561 AsnTrpThr.....ProLYAsnTYrValProH1sProGluArG1 574
    ||| ..... |||
4311 AATAAACCAGACGAATGCTAAATAACTAT...CCAAGCCGAGAAATAT 4357
    ||| ..... |||
574 nSerThrLeu.ValAlaAsnThrLeuTrpAsnThrTYrSerAsp.MetC1 590
    ||| ..... |||
4358 TACTACTCAACGTAGTAAACGCGCTTGG.....GACCAATGCA 4395
    ||| ..... |||
590 nAlaValGlnSerMet.ILeasnThrIleAlaH1sG1YAlaTYrLeu 606
    ||| ..... |||
4396 GACGTGTCAACGTAGCTTAAACCATTAACCTGTTCACACACACTGTATC 4445
    ||| ..... |||
607 PhG1YThrTrpG1YSerAlaValSerAsnLeuPhetYrAlaH1sAspse 623
    ||| ..... |||
4446 ATT..... 4448
    ||| ..... |||
623 rSerG1YsFroIleaspAsnTrpH1sH1sArgSerLeuG1YTrLeup 640
    ||| ..... |||
4449 .....AAACAGATAGAC.....CACAGCAGAGAAATGAAGATTC. 4484
    ||| ..... |||
640 heG1YIleSerThH1sSerLeuAspH1sSerPhG1YLeuAla1a 656
    ||| ..... |||
4485 .....ATTACTAAATCTTCCTATACAGATCTCTCATTTAAAGCCGC 4529
    ||| ..... |||
657 GlycInleuLeuG1YsSerSerAspserPhelIethrSerThG1uTh 673
    ||| ..... |||
4530 ACTCAGACA.....AATAGTTCTGACTACACATTTGTAACATCAGC 4573
    ||| ..... |||
673 rThrSerTrIleAlaThrValGlnAlaGlnLeuAlaThrSerLeuMetL 690
    ||| ..... |||
4574 ATTTCACCAATAGATCTCAAAAGAGATTCATTCAAACAAATTTCTC 4623
    ||| ..... |||
690 ySlIeSerAlaGlnAlaCysTYrAsnGluSerIleH1sG1YLeuLysThr 706
    ||| ..... |||
4624 ATGTTCAGACA.....TCATCTACATTATAGACTTTAAAGACA 4661
    ||| ..... |||
707 Lys 707
    |||
4662 AAA 4664

```

seq_name: /cgn2.6/prodata/2/ina/PCTUS_COMB.seq:PCT-US94-13200-5

```

seq_documentation_block:
: Sequence 5, Application PC/TUS9413200
: GENERAL INFORMATION:
: APPLICANT: Emory University
: TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Kilpatrick & Cody
: STREET: 1100 Peachtree Street, Suite 2800
: CITY: Atlanta
: STATE: Georgia

```

```

: COUNTRY: US
: ZIP: 30309
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/13200
: FILING DATE: 15-NOV-1994
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Padst, Patrea L.
: REGISTRATION NUMBER: 31,284
: REFERENCE/DOCKET NUMBER: EMU106CIP(2)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 404-815-6508
: TELEFAX: 404-815-6555
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7493 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Mus musculus
: FEATURE:
: NAME/KEY: repeat_unit
: LOCATION: 1..407
: OTHER INFORMATION: /rpt_type="terminal"
: OTHER INFORMATION: /note="5'UTR"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 7471..7476
: OTHER INFORMATION: /function="PolyA_signal"
: FEATURE:
: NAME/KEY: repeat_unit
: LOCATION: 7368..7493
: OTHER INFORMATION: /rpt_type="terminal"
: OTHER INFORMATION: /note="3'UTR"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 408..7367
: OTHER INFORMATION: /product="Coagulation Factor VIII"
: PUBLICATION INFORMATION:
: AUTHORS: Elder, F.
: AUTHORS: Lakich, D.
: AUTHORS: Gletschler, J.
: TITLE: Sequence of the Murine Factor VIII cDNA.
: JOURNAL: Genomics
: VOLUME: 16
: PAGES: 374-379
: DATE: 1993
: RELEVANT RESIDUES IN SEQ ID NO: 5: FROM 1 TO 7476
: PCT-US94-13200-5

```

```

alignment_scores:
: Quality: 141.00 Length: 753
: Ratio: 0.387 Gaps: 39
: Percent Similarity: 48.340 Percent Identity: 21.248

```

alignment_block:

US-09-677-752-4 x PCT-US94-13200-5 ..

```

Align seg 1/1 to: PCT-US94-13200-5 from: 1 to: 7493
46 PhG1YAspCysThr.....MetGluSerLeuPhetProAl 57
||| ..... |||
2730 TTCAAGATTCCACAAATTCCTCAAAATATGATATGAGAGATTGACCTCA 2779

```

```

3900 TTTGTCAGAGAAAAATTAAGTCAGACAGATGATGATTTTCAAGAA 3949
448 uilewcluiagiThrSerLeuLysThrSerSeraspLeuLysLeuA 465
3950 CATAGAGACTTAAGACATGCGCTTTCCACATATATGACCATTTCTTA 3999
465 laThrLeuSerLleProLeuHisSer.....LeuAspThrGluLys 478
4000 CCACCTTGTCT...AACGTACATGAAATGTAGCAGCATCAAGAAAAA 4046
479 SerVal.....ThrLleHisAlaProAsnLeuSerLleGlnLysLlePh 493
4047 AATATTCAAGAAAGATAGAACAGAACACTAAATTGAAGAAAGTAGT 4096
493 eleuSerasn.....SerGlyAspGluAsnPhetyrGluAsnV 506
4097 TTTGCCCCAGGTGCGACAGACAGACAGCTGCTTACATTTCTTGAAGACA 4146
506 aGluLeuLeuSerLysGluGlnAsn..... 514
4147 TATTGATCTAGGACCTAGCGCAAAATATAAGTTATATGAGTACATGTA 4196
515 .....AsnLleProLeuLeuThrLeuSerLysGluGlnSerHi 527
4197 CCAGTACTTCAAAACATCATCATTAACAATTCACAAATACAGTACA 4246
527 sleuHisLeuProAspGlyAsnLeuSerSerHisphGlytyrGlnGlyA 544
4247 CATTGACATG.....GACCATTTCTTTTAAACAAGC. 4277
544 spTrpThrPhSerTrpLysAspSerAspGluGlnHisSerLeuLleA 560
4278 .....AAGGCAAGAAACAAATTCAGAAAGCGCTGGTA 4310
561 AsnTrpThr.....ProLysAsnTyValProHisProGluArgG 574
4311 AATAAAGACAGAGAAATGTAATAAAATAT...CCAACCAAGCAAGATAT 4357
574 nserThrLeuValAlaAsnThrLeuTrpAsnThrTySerAsp_MetG 590
4358 TACTAGCTCAAGCTAGTAAACGGCGCTTTCG.....GACATTTCA 4395
590 naLalaGlnSerMet_LleasnThrLleAlaHisGlyGlyAlaTyLeu 606
4396 GACTGTCAACTCAATGCTTAAACCATTAACCTGTTCACACAGTATAC 4445
607 pheGlyThrTrpGlySerAlaValSerAsnLeuPhetyrAlaHisaspse 623
4446 ATT..... 4448
623 rSerGlyLysProLleAspAsnTrpHisAspSerLeuGlyTyLeuP 640
4449 .....AACAGATAGAC.....CACAGCAAGAAATGAAAAAGTTTC. 4484
640 heGlyLleSerThrHisSerLeuAspAspHisSerPheCysLeuAlaA 656
4485 .....ATTACTAAATCTCTCTATCAGATCTCTGTGATTAAGAGCAC 4529
657 GlyGlnLeuLeuGlyLysSerSerAspSerPheLleThrSerThGluTh 673
4530 ACTCGACAA.....AATAGTTCCTACCTCACACATTTGAAAAACATCGAC 4573
673 rThSerTyrlleAlaThrValGlnAlaGlnLeuAlaThrSerLeuMetL 690
4574 ATTTCGACCATAGATCTCAAAAGACATTCACAAACAAATTTTTCGC 4623
690 yslleSerAlaGlnAlaCysTyraSngLysLleHisGluLeuLysThr 706
4624 ATGTTCAAGCA.....TCATCTACATTTATGACTTTAGACGA 4661
707 Lys 707

```

```

4662 AAA 4664
seq_name: /cgn2_5/prodata/2/lna/5b_COMB.seq:us-09-037-601-5
seq_documentation_block:
? Sequence 5, Application US/09037601
? Patent No. 6180371
? GENERAL INFORMATION:
? APPLICANT: Lollar, John S.
? TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
? NUMBER OF SEQUENCES: 40
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
? STREET: 5370 Manhattan Circle Suite 201
? CITY: Boulder
? STATE: Colorado
? COUNTRY: USA
? ZIP: 80303
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/037,601
? FILING DATE: 26-JUN-1996
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: NO PCT/US94/13200
? FILING DATE: 15-NOV-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/212,133
? FILING DATE: 11-MAR-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/864,004
? FILING DATE: 07-APR-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Ferber, Donna M.
? REGISTRATION NUMBER: 33,878
? REFERENCE/DOCKET NUMBER: 75-95F
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 303/499-8080
? TELEFAX: 303/499-8089
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 7493 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: not relevant
? MOLECULE TYPE: cDNA to mRNA
? HYPOTHETICAL: NO
? ANTI-SENSE: NO
? ORIGINAL SOURCE:
? ORGANISM: Mus musculus
? FEATURE:
? NAME/KEY: repeat_unit
? LOCATION: 1..407
? OTHER INFORMATION: /rpt_type="terminal"
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 7471..7476
? OTHER INFORMATION: /function="polyA signal"
? FEATURE:
? NAME/KEY: repeat_unit
? LOCATION: 7368..7493
? OTHER INFORMATION: /rpt_type="terminal"
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 408..7367
? OTHER INFORMATION: /product="coagulation factor VIII"
? PUBLICATION INFORMATION:

```

```

3763 AGAGTGTCTCTGTCA.....GAATCTTCAAAAT 3791
385 LeuValLLegInThrSerAlaAsnArgSerValIlePheSerSerClar 401
3792 TGGTTAAAGAACCAATGAAATATATCTTGG..... 3824
401 gLeuSerGIuGIuGIuThrProAspAsnLeuThrSerGIuLeuGIu 418
3825 AACCTGACCAAGAACATAGTCCAAACCAATAGTATATTATTAATG.... 3869
418 InProIleGIuLeuSerGIuArgLeuValIleuLysAspArgAlaVal 434
3870 .....TTTAAAAATATGTAAAAATCAAAAGTTTC 3899
435 LeuSerAlaPro.....SerLeuSerGIuAspProGIuAlaLeuLe 448
3900 TTGTTCAGAGAAAAATAAGTCACATGACATGACATGATTTACAAAGAA 3949
448 uIleuGIuAlaGIuThrSerLeuLysThrSerSerAspLeuLysLeuA 465
3950 CATAGGACTTAAGACATGGCTTTTCACATATATGACATATTTCTTA 3999
465 LeuThrLeuSerIleProLeuThrSer.....LeuAspThrGIuLys 478
4000 CCACCTTGTCT...AACGTACATGAAATGTAGGCACATCAAGAAAAA 4046
479 SerVal.....ThIleIleAlaProAsnLeuSerIleGIuLysIlePh 493
4047 AATATTCAGACAGACATGACACAGACACTAATGACAGAAAGTACTGT 4096
493 eLeuSerAsn.....SerGIuAspGIuAsnPheTyrgIuAsnV 506
4097 TTTCCCCAGCTCCACAGAACCAACGCTCTTACGAATTTCTTGAAGACA 4146
506 aIGIuLeuLeuSerLysGIuGIuAsn..... 514
4147 TATTGATCTAGGCACTAGGCAAAATATTACTTATATGAGTACATGTA 4196
515 .....AsnIleProLeuLeuThrLeuSerLysGIuGIuSerHI 527
4197 CCAGTACTTCAAAACATCATCATATACAAATTCACAAATACAGTACA 4246
527 sLeuHIleuProAspGIuAsnLeuSerSerHIlePheGIuTyrgIuGIu 544
4247 GATTCACATG.....GACCATTCCTTTAAAGAACAGG. 4277
544 sPTPrThrPheSerTrpLysAspSerAspGIuGIuHIleSerLeuIleA 560
4278 .....AAGGACAAAGGAAACAAATTCGAAGGCTTGTA 4310
561 AsnTrpThr.....ProLysAsnTyrgAlProHIspProGIuArgI 574
4311 AATAAACCAAGAAATGCTAAAAACTAT...CCAAGCCAGAACATAT 4357
574 nSerThrLeuValAlaAsnThrLeuTrpAsnThrTyrgSerAspMetGI 590
4358 TACTACTCAACCTAGTAAACGCGCTTGG.....GACAAATCA 4395
590 nAlaValGIuSerMetIleAsnThrIleAlaHIleGIuAlaValTyrgLeu 606
4396 GACGTGCACTCAATGCTTAAACCAATAACTGTTCACACAGTGTATC 4445
607 PheGIuThrTrpGIuSerAlaValSerAsnLeuPheTyrgAlaHIleAspSe 623
4446 ATT..... 4448
623 rSerGIuLysProIleAspAsnTrpHIleHIleArgSerLeuGIuTyrgLeuP 640
4449 .....AAACAGATAGAC.....CACACCAAGCAAAATGCAAAAGTTC. 4484
640 heGIuLysSerThrHIleSerLeuAspAsnHIleSerPheCysLeuAlaIleA 656
4485 .....ATTACTAAATCTTCTTATCATGATCTCTGTGATTAAGACAC 4529

```

```

657 GYGIuLeuLeuGIuLysSerSerAspPheIleThrSerThrGIuTh 673
4530 ACTCAGACA.....AATAGTTCTGACTCAGACACATGTAACCAATCAGC 4573
673 rThrSerTyrgIleAlaThrValGIuAlaGIuLeuAlaThrSerLeuMetI 690
4574 ATTTCACATCATATAGATCTCAAAAGAGTCATTCCAAAACAATTTTCTC 4623
690 ySIleSerAlaGIuAlaCysTyrgAsnGIuSerIleHIleGIuLeuLysThr 706
4624 ATGTTCAAGCA.....TCATCTACATTTATGACTTAAAGACA 4661
707 Lys 707
4662 AAA 4664

seq_name: /cgn2_6/prodata/2/lna/5B_COMB.seq:us-08-670-707A-5
seq_documentation_block:
; Sequence 5: Application US/08670707A
; Patent No. 5859204
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/670,707A
; FILING DATE: 26-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/13200
; FILING DATE: 15-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,133
; FILING DATE: 11-MAR-1994
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Greenlee, Lorraine L.
; REGISTRATION NUMBER: 27,894
; REFERENCE/DOCKET NUMBER: 75-95F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; INFORMATION FOR SEQ. ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: repeat_unit
; LOCATION: 1..407
; OTHER INFORMATION: /rpl_type= "terminal"

```

```

294 AlaIle.....AlaIleGlnSerGlySerLeuSerI 305
|||||
3561 GCATATTAAAGTCATATGATGATCAA.....GAGTTAACGC 3601
305 eLeuAlaGlyGlnSerValLeuPheGlnAsnSerGlnThrS 322
|||
3602 TTTCATTCATGATGACACCTTTTAAAGCAAAATTCACA..... 3641
322 eAspGlnGlyLeuValAlaGlnAlaIleTyrLeuGlnLysAlaIle 338
|||||
3642 .....TATTGACACTAACCATATG 3662
339 LeuSer...SerLeuGlnAlaArgAsnGlyAspIleLeuPhe..... 351
|||||
3663 CTAATATAGACTACTCAACAAATAAGACATATTTCATAGAAAGA 3712
352 .PheAspProIleValGlnGlnSerSerLysGlnSerProLeuProS 368
|||||
3713 TCAACATCTCTTCCACAAGATGAGAAATACATCATGCCATTTCACA 3762
368 eSerLeuGlnAlaSerValThrSerProThrProAlaThrAlaSerPro 384
|||||
3763 AGATCTTCTTCTGTCA.....CAATCTTCAAAW 3791
385 LeuValIleGlnThrSerAlaAsnArgSerValIlePheSerSerGluAr 401
|||||
3792 TGCTTAAAGACCAATGGAATATTCCTTG..... 3824
401 gLeuSerGlnGluGlyThrProAspAsnLeuThrSerGlnLeuGlnG 418
|||||
3825 .AACCTGACGAGAACATAGTCCAAACCAATGATATTTTAAAG.... 3869
418 InProIleGlnLeuSerGlyArgLeuValIleLysAspArgAlaVal 434
3870 .....TTTAAAAATATGTATAAAATCAAGCTTTC 3899
435 LeuSerAlaPro.....SerLeuSerGlnAspProGlnAlaLeuLe 448
|||||
3900 TTGTACAGAGAAAATTAAGTCCAGTAGACAGATGATTACAAAGAA 3949
448 uIleMetGlnAlaGlyThrSerLeuLysThrSerSerAspLeuLysIleuA 465
|||||
3950 CATAGACTTAAAGACATCCCTTTTCCACATATATGAGCAATTTTCCTAA 3999
465 laThrLeuSerIleProLeuIleSer.....LeuAspThrGlnLys 478
|||||
4000 CCACCTTTGCTT...AACGTACATGAATATGAGCACAATCAAGAAAAA 4046
479 SerVal.....ThrIleHisAlaProAsnLeuSerIleGlnLysIlePhe 493
|||||
4047 AATATTTCAGGAAGATAGAGAAAGACACTAATTGAAGAAAGAGTGT 4096
493 eLeuSerAsn.....SerGlyAspGlnAsnPheTyrGlnLysAsnV 506
|||||
4097 TTTCGCCAGGTCACAGAGCAACCTGCTTAAGAAATTTCTTGAAGACA 4146
506 aIleGlnLeuSerLysGlnGlnAsn..... 514
4147 TATTATATCTAGGCACTAGCCAAATATATAGCTTTATATGACTACATGTA 4196
515 .....AsnIleProLeuLeuThrLeuSerLysGlnGlnSerIle 527
|||||
4197 CCAGTACTTCAAAACATCATCATTAACATTAACAAATACAGTACA 4246
527 sLeuHisLeuProAspGlyAsnLeuSerSerHisPheGlyTyrGlnGlyA 544
|||||
4247 GATTACACATG.....GAGCATTTCTTTAAACAAGC. 4277
544 sPTrpThrPheSerTrpLysAspSerAspGlnGlyHisSerLeuIleAla 560
|||||
4278 .....AAGGACAAGAAACAATTCAGACAGCTTTGTA 4310

```

```

561 AsnTrpThr.....ProLysAsnTyrValProHisProGlnAlaArgI 574
|||||
4311 AATAAACCAGAGAAATGTAAAAACATAT...CCAAACGACAGAAATAT 4357
574 nSerThrLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMetG 590
|||||
4358 TACTACTCAACGTAGTAAACGGCTTTG.....GACAAATTC 4395
590 nAlaValGlnSerMetIleAsnThrIleAlaHisGlyValTyrLeu 606
|||||
4396 GACTGTCAACGTCAATGCTTAAACCATTAACGTGTCAACACAGTGTATC 4445
607 PheGlyThrTrpGlySerAlaValSerAsnLeuPheTyrAlaHisAspSe 623
4446 ATT..... 4448
623 rSerGlyLysProIleAspAsnTrpHisHisArgSerLeuGlyTyrLeuP 640
|||||
4449 .....AACACATAGAC.....CACAGCAGCAAAATGAAAAAGCTTC. 4484
640 nGlyIleSerThrHisSerLeuAspAspHisSerPheCysLeuAlaIa 656
|||||
4485 .....ATTACTAAATCTTCTCTTATCAGATTCTCTGTGATTTAAAGCACC 4529
657 GlyGlnLeuLeuGlyLysSerSerAspSerPheIleThrSerThrGluTh 673
|||||
4530 ACTCAGACA.....AATAGTCTGACTCACACATTTGTAACATCAGC 4573
673 rThrSerTyrIleAlaThrValGlnAlaGlnLeuAlaThrSerLeuMetL 690
|||||
4574 ATTTCACCAATAGATGATCTCAAAAGAGTCATTCCAAAACAATTTCTC 4623
690 yIleSerAlaGlnAlaCysTyrAsnGlnSerIleHisGlnLeuLysThr 706
4624 ATGTTCAAGCA.....TCATCTACATTTATGACTTTAAGACA 4661
707 Lys 707
4662 AAA 4664

```

seq_name: /cgn2_6/prodata/2/lna/5A_COMB.seq:US-08-474-503-5

```

seq_documentation_block:
: Sequence No. 5744446
: Patent No. 5744446
: GENERAL INFORMATION:
: APPLICANT: Emory University
: TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Kilpatrick & Cody
: STREET: 1100 Peachtree Street, Suite 2800
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: US
: ZIP: 30309
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/474,503
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Pratt, John S.
: REGISTRATION NUMBER: 29,476
: REFERENCE/DOCKET NUMBER: EMU106CIP(3)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 404-815-6500
: TELEFAX: 404-815-6555
: INFORMATION FOR SEQ ID NO: 5:

```


284 pheleasnanseraleuylsileuylalaleuylleuylsergi 300
1886 ACTGGCGGCTCATATCATATATACCGGAATCTTACTGTTCAAAAG 1935
300 ygliserleuylleuylalagly 312
1936 CCTAACCTTCAAGCTATACAAATTTACACTTTATATAGCCGCTCA 1984
312 alleuphegluasnserser 318
1985 TTTGACACACATAGCGCTTCAACATTTCCATTGCCAGAGAGG 2029
319 Glnatphtseraspngllyleuvalargas 329
2030 GCTAATTTAAAGATATCATATACACAGTAGCTTAATATATACACCA 2079
329 nalaleuylleuylleuylaspaalaleuylserleuylalarg 346
2080 CTCTGATACCACTTACCCACCATATATAAGCAATATATCCACCAAT 2129
346 snglyaspheleupheasp 355
2130 CAGGTGATTTGATATATATGATATAAAGCGACGCTGAATCCAAAT 2179
356 valcngluserserleuylserleuylproleuylserleuyl 371
2180 GCGGCAATATCTCTCAAAAGAGCAATCTCACATTTCTCTGATATA 2229
371 nalaserleuylserleuylproleuylleuylaserleuylval 388
2230 AGTAATATTTACCAAT 2246
388 lnthtserleuylaspaalaleuylserleuylargleuylsergi 404
2247 AGATACATCATACAGCGCTTCAAGGGGCGCTTGTGATTCAGATG 2296
405 glugllylthrtproaspnleuylserleuylleuylleuylleuyl 421
2297 GCAGAAATGCT 2331
421 uleuyls 2381
2332 GTTAAATTTGGCAGCAGCTTAATTTTCAAGCTTATATAACCAAA 2381
435 euserleuylproleuylserleuylproleuylleuylleuylleuyl 451
2382 TTACGCTTAAATGCGAGTAT 2410
452 Alaqlthrtserleuylthrtserleuylleuylleuylleuyl 466
2411 ATTGGCATGCTAGCGGTGATATGCTGATATAAAGTACTTTGA 2460
467 2510
2461 CAAGCTTAACATTCATAAATCTCCACGCTACCGCTCAATTAACACAA 2510
477 lulyserleuylthrtleuylleuylleuylleuylleuylleuyl 493
2511 ATAGGGAAGT 2539
494 leuylserleuylleuylleuylleuylleuylleuylleuylleuyl 510
2540 ACCATGCTGCTAATGATATAC 2583
510 rlysglu 524
2584 AAAACATTAACGCTTAACAAATTAACGCT 2621
524 luginserleuylleuylleuylleuylleuylleuylleuylleuyl 540
2622 AGACATATAATCTCTGCGGAGCAGAAATGTACAACCAAAAGAGC 2671
541 tyrgln 556

2672 ACAATATCAATGCAACCAAGCCGCTGGAAGTAATGCTCAAAATG 2721
556 sserleuylleuylleuylleuylleuylleuylleuylleuylleuyl 573
2722 TACAATTAAGCAACATTAACCTCCGCAAAATGTACAGTACAGCAAG 2771
573 rglserleuylleuylleuylleuylleuylleuylleuylleuylleuyl 589
2772 AAATCTGTTTACACAGAGATGCTGATTAATGACCAAGCGGCA 2821
590 Glnalvalglnserleuylleuylleuylleuylleuylleuylleuyl 606
2822 GAAACATTAAGCAACCAAGCGGCTGATTAAGTGA 2862
606 uphegllylthrtgly 619
2863 TCAATCAACCTCCGCTAATGTAATATTAACAGCGGCAATACATTA 2912
619 yralanisaspserleuylleuylleuylleuylleuylleuylleuyl 635
2913 AGGTAGTAATATCACTGCTCAAGTGA 2941
636 leuyllylthrtleuylleuylleuylleuylleuylleuylleuylleuyl 2942
652 ecylleuylleuylleuylleuylleuylleuylleuylleuylleuyl 669
2968 GACACATACAGCAGCTCAACCATTAAGTGCAGCAACAGCAATGCAATA 3017
669 hrserleuylleuylleuylleuylleuylleuylleuylleuylleuyl 684
3018 TTACACCAAAACAGCTATATCAACGTAATGTAATTCACCTCCGCG 3067
685 Alathrtserleuylleuylleuylleuylleuylleuylleuylleuyl 701
3068 TCTGTAACACTGTTGTAACAGTACAGCACTCTGCTGA 3106
701 ehlsleuylleuylleuylleuylleuylleuylleuylleuylleuyl 718
3107 3116
718 ertrphtserleuylleuylleuylleuylleuylleuylleuylleuyl 731
3117 CAGTAACACTGTTACTATTAAGCGGATAGCGGAATTAACCTCCCA 3166
732 lileuylleuylleuylleuylleuylleuylleuylleuylleuylleuyl 748
3167 GTAGTCTTACAAATTAAGGCTATATAGTGAACACCTCAAGCAATC 3216
748 eserleuylleuylleuylleuylleuylleuylleuylleuylleuyl 764
3217 AGCGATATTAAGGTACAAATTTGTAATTAACAGTAATGTTACAGCA 3266
764 erserleuylleuylleuylleuylleuylleuylleuylleuylleuyl 768
3267 GCACGTGATTTA 3280

seq_name: /cgn2.6/prodata/2/ina/5a_COMB.seq:us-08-212-133A-7
seq_documentation_block:
; Sequence 7, Application US/08212133A
; Patent No. 5663060
; GENERAL INFORMATION:
; APPLICANT: Lohlar, John S.
; APPLICANT: Runge, Marschall S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 100 Peachtree Street
; CITY: Atlanta

170 SerSerLysGlyAsnGlyAlaIleGlnAlaGlnThrPheSerLeuSe 186
1448 ATTGAGTCTAGGGGGGGGTAGCTTAATTCAAACTTAACGCTCATC 1497
186 rArgAsnValSerPro.....IleSerPheAlaArgAsnArgAlaAspL 201
1498 TAGCAACATACAAACCCTGGCGTAATTATAAATCTCAAAACCTTAATG 1547
201 euAsnGlyAlaIleGlyCysSerAsnLeuIleCysSerGlyAsnVal 217
1548 TCTCAGAGGGTCAACTTA.....AATCTCAAGGCTGAAGGTTCAACA 1591
218 AsnProLeuPhe.....PheThrGlyAs 225
1592 GAACCCGCTTTTCAATAGAAAATGATTTAAACTTAACGCCACCGGTGG 1641
225 nSerAlaThr..... 228
1642 CAATTAACAATCAGACAGTCGAGGTACCGATTACGCCGTCAACAAAG 1691
229AsnGlyAlaIleGlyCys 235
1692 GTGTGGCAGCCAAAAAACATACCTTTAAAGGGGTAAATTCACCTTC 1741
236 IleSerAspLeuAsnThrSerGlu...LysGlySerLeuSerLeuAlaCyl 251
1742 GGCTCTCAAAAAGCCACAAACAAATCAAGGCATGTTACCATCAATAA 1791
251 sAsn...GlnGluThrLeuPheAlaSerAsnSerAlaLysGluLysGlyG 267
1792 AAMCCTAACCTACTCTTCGTGGTGCAGATTTTCCGCAAAACAAATCGC 1841
267 LyAlaIleThrAlaLysHisMetValLeuArgTyraAsnGlyProValSer 283
1842 CTTTAATATATAGCAGGAATGTTATT.....AATAATGGCACCTTACC 1885
284 PheIleAsnAsnSerAlaLysIleGlyAlaIleAlaIleGlnSerGly 300
1886 ACTGCCGCTCATATCAATATACCGCAAAATCTTACTGTTCAAAAG 1935
300 yGlySerLeuSerIleLeuAlaGly.....GluGlySerV 312
1936 CGCTAACCTTCAAGCTATACAAATTAACCTTTAATGTAGCCGCTCA. 1984
312 AlLeuPheGlnAsnAsnSer..... 318
1985TTTGACACAAATGGCGCTTCAACATTTCCATGCCAGAGAGAG 2029
319GlnArgThrSerAspGlnGlyLeuValArgAs 329
2030 GCTAATATTAAGATATCAATATACACAGTACCTTAATATATACACCA 2079
329 nAlaIleThrLeuGluLysAspAlaIleLeuSerSerLeuGluAlaArgA 346
2080 CTCGTATACCACTTACCGCACCATTTATAAAGCAATATATCCAACAAT 2129
346 sngIAspIleLeuPhePheAsp.....ProIle 355
2130 CAGGCGATTGATATATATGATATAAAAAAGCGCTGAATTCGAAT 2179
356 ValGlnGluSerSerSerLysGluSerProLeuPro...SerSerLeuGly 371
2180 GCGCGCAATATCTCAAAAAGCAAGCAATCTCACAAATTTCTTCATGA 2229
371 nAlaSerValThrSerProThrProAlaThrAlaSerProLeuValIleG 388
2230 AGTAATATTAACCAAT.....C 2246
388 lnthSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerGlu 404
2247 AGATAACAATCAACAGCGCTTGAAGGGGGCTTCTGATTCAGAGTGAG 2296

405 GluGluLysThrProAspAsnLeuThrSerGlnLeuGlnProIleGly 421
2297 GCAGAAATGCT.....AACCTAATCTATTCACCAACAA.....GA 2331
421 uLeuLys...SerGlyArgLeuValLeuLys.....AspArgAlaValL 435
2332 GTTAAATTTGGCAGAGACCTAAATTTTCAAGGCTTTAATTAACCGAAA 2381
435 euSerAlaProSerLeuSerGlnAsnProGlnAlaLeuLeuIleMetGlu 451
2382 TTACAGCTAAATAATGCCAGTCAAT.....TTAACT 2410
452 AlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAlaThr.... 466
2411 ATTGCATCTGAGCGGTGATGATGCTGATTAATAAAGAGACTTTTGA 2460
467LeuSerIleProLeuHisSerLeuAspThrG 477
2461 CAAGGTTAAGATTCCAAAATCTCGACTGAGCGGTGCACATATTAACCTA 2510
477 LysSerValThrIleHisAlaProAsnLeuSerIleGlnLysIlePhe 493
2511 ATACCGAAGTC.....AAACGCTTAATGCTACT..... 2539
494 LeuSerAsnSerGlyAspGluAsnPheTyrgluAsnValGluLeuLeuSe 510
2540 ...ACCAATGCGGTGAATGATAAC...AGCACCGGTTAAACCATTTCCGC 2583
510 rLysGlu.....GlnAsnAsnIleProLeuLeuThrLeuSerLysG 524
2584 AAMAGATTAACGGTAACAAATAACGTT.....ACCTCCGCA 2621
524 LuGlnSerHisLeuHisLeuProAspGlyAsnLeuSerSerHisPheGly 540
2622 AGACATTAATATCTCTCCCGCAGCAGGAATGTACACCAACAAAGAGC 2671
541 Tyrgln...GlyAspThrPheSerThrPheLysAspSerAspGluGlyH 556
2672 ACAACTATCAATGCACACACAGCGCGTGAAGTAAGTGCCTCAAAATG 2721
556 sSerLeuIleAlaAsnThrPheProLysAsnTyraProHisProGluA 573
2722 TACAATTAAGGCACATTAACCTCGCAAAATGTACAGTGCAGCAACAG 2771
573 rGlnSerThrLeuValAlaAsnThrLeuThrPasnThrTyraSerAspMet 589
2772 AAAATCTGTATACCAACAGAAATGCTCATTAATGCAACACCGCGCA 2821
590 GlnAlaValGlnSerMetIleAsnThrIleAlaHisGlyAlaTyraLe 606
2822 GTAAACATTAGACAAAAACAGCGGATATTAAGGTGGA.....AT 2862
606 uPheGlyThrTrpGly.....SerAlaValSerAsnLeuPheT 619
2863 TGAACTCAACTTCGGTAATGTAATATTAACAGCGGCAATACACTTA 2912
619 yAlaHisAspSerSerGlyLysProIleAspAsnThrHisArgSer 635
2913 AGCTAGATTAATATCTGTCAGATGTA..... 2941
636 LeuGlyTyraLeuPheGlyIleSerThrHisSerLeuAspAspHisSerPh 652
2942ACAGTACACCGGATGACGAGGCTT 2967
652 ecysLeuAlaIleGluGlnLeuLeuGlyLysSerSerAspPheIleT 669
2968 GACAACTACACGAGGCTCAACCATTAAGTGCACACAGCAAGCAATA 3017
669 hrSerThrGluThrSerTyraIleAlaThrValGlnAlaGlnLeu... 684
3018 TTACAAACCAAAACAGCTATATCAACGGTAAAGTTGAATCCACGCTCGGC 3067
685 AlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyraAsnGluSerI 701

MOLECULE TYPE: DNA (genomic)
US-08-530-198-7

alignment_scores:

Quality: 144.00 Length: 788
Ratio: 0.364 Gaps: 38
Percent Similarity: 50.254 Percent Identity: 19.670

alignment_block:
US-09-677-752-4 x US-08-530-198-7

Align seg 1/1 to: US-08-530-198-7 from: 1 to: 4287

```
72  GlysAsnSerTyrCysTTrpPheValSerLysLeuHisIleThrAspProLys 88
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1208 GGAGCGACACTAC...TGGACGTAACACCTTAATGTTACCTCG... 1249
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88  sGluAlaLeuPheLysGluLysGlyAspLeuSerIleGlnAsnPheArgP 105
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1250 .....CGTAAGTAATTTAAACCTCCATTCACAGC..... 1279
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
105  heLeuSerPheThrAspCysSerLysGlyIleSerSerProSerIleIle 121
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1280 .....ACAGGAAGTGGCTCAACAGGCTCAAGCATA... 1309
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122  HisGlnLysAsnGlyLysIleLeuSerLeuArgAsnGlySerMetSerPh 138
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1310 .....CCCATTCGACGATTAATGCCATTAACATTAAATTAAGCCACTT 1353
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
138  eCysArgAsnHisAlaGlyLysGlyLysGlyLysAlaIleSerAla 153
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1334 T.....AATATCGCACACCGCTCAACAGCTAATTATTAATGAT 1397
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153  sPalaPheSerLeuGlnHisAsnTyrIlePheThrAlaPheGlnGluAsn 169
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1398 CAATATATCCCTTTAAGAGTAACGCTAAGCTATTAATTAATGAT 1447
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
170  SerSerLysGlyAsnGlyLysAlaIleGlnIleGlnIlePheSerLeu 186
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1448 ATTTCAGTCTCAGCGCGGCTAGCTTAATTTCAACACTTAACGCCCTCATC 1497
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
186  rArgAsnValSerPro.....IleSerPheAlaArgAsnArgAlaAspL 201
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1498 TACCAACATACAAACCCCTGGCTGATTAATTAATCAAAACCTTAATG 1547
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201  euAsnGlyAlaIleLecysCysSerAsnLeuIleCysSerGlyAsnVal 217
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1548 TCTCAGGAGGCTCACTTA.....AATCTCAAGGCTGAAGTTCAACA 1591
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218  AsnProLeuPhe.....PheThrGlyAs 225
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1592 GAAACCGCTTTTCAATAGAAATGATTTAAACCTTAACGCCACCGCTG 1641
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
225  nSerAlaThr..... 228
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1642 CAATATACCAATCAGACAGTGCAGCTACGATTCACGCCGTCACAAG 1691
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
229 .....AsnGlyAlaIleLecys 235
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1692 GTGTGCGACCAAAAAACATACCTTTAAAGGGGTAATATACCTTC 1741
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
236  IleSerAspLeuAsnThrSerLysLysGlySerLeuSerLeuAlaLys 251
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1742 CGCTCTCAAAAAACCAACAGAAATCAAGCATGTATACCATCAATTA 1791
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251  sAsn...GlnGluThrLeuPheAlaSerAsnSerAlaLysGlyLysGlyG 267
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1792 AAACACTACGCTACTCTTCGTGGTGCAGAAATTTCCGCAAAACAAATCC 1841
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
267  yAlaIleTyrAlaLysHisMetValLeuArgTyrAsnGlyProValSer 283
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1842 CTTTAATATATACAGAAATGTATT.....AATATGCGACACTTAC 1885
```

```
284  PheIleAsnAsnSerAlaLysIleGlyAlaIleAlaIleGlnSerG 300
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1886 ACTGGCGGCTCCATTAATCAATATAGCCGGAATCTTACTGTTCAAAAG 1935
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
300  yGlySerLeuSerIleLeuAlaGly.....GluGlySerV 312
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1936 CGTAACCTTCACGCTAATACAAATTAACACTTTAATGTTACCGCTCA 1984
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
312  aLeuPheGlnAsnAsnSer..... 318
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1985 .....TTTGACACAAATGCGCTTCAACATTTCCATTGCCAGAGAGG 2029
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
319 .....GlnArgThrSerAspGlnGlyLeuValArgAs 329
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2030 CCTAATTTAAGATATCAATACACCGCTAGCTTAATATTAACACCA 2079
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
329  nAlaIleTyrLeuGlnLysAspAlaIleLeuSerSerLeuGluAlaArg 346
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2080 CTCTGATACCACTTACCGCACCATTTAAGCAATATATCAACAAAT 2129
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
346  snglyAspIleLeuPhePheAsp.....ProIle 355
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2130 CAGTGATTTGAATATTTATGATTAATAAAACGACGCTGAATCCAAAT 2179
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
356  ValGlnGluSerSerSerLysGlySerProLeuPro...SerSerLeuG 371
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2180 GCGCGCATATCTCCACAAAGAGGCAATCTCACAATTTCTTGATGA 2229
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
371  nAlaSerValThrSerProThrProAlaThrAlaSerProLeuValIleG 388
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2230 AGTAATATTTACCAAT.....C 2246
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
388  lInThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerG 404
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2247 AGATACCAATCAAGCAGCGCTGAAGGGGCGCTTGATTAAGTAG 2296
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
405  GluGluLysThrProAspAsnLeuThrSerGlnLeuGlnGlnProIleG 421
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2297 GCAGAAATGCT.....AACCTAATCTATTCACAAACCA.....GA 2331
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
421  uLeuLys...SerGlyArgLeuValLeuLys.....AspArgAlaValL 435
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2332 GTTAAATATGACAGAGACCTAATATTTCAAGGCTTTAATAAGCAGAA 2381
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
435  euSerAlaProSerLeuSerGlnAspProGlnAlaLeuLeuIleMetGlu 451
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2382 TTACAGCTAAATAATGCGAGTAT.....TTAAGT 2410
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
452  AlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAlaThr.... 466
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2411 ATTGCATATGCTAGCGGTGTAATGCTGATGCTAATAAAGTACCTTTGA 2460
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
467 .....LeuSerIleProLeuHisSerLeuAspThrG 477
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2461 CAAAGTTAAAGATTCAAAATCTCGACTGACGGTCAACATGTAACACTTA 2510
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
477  LuLysSerValThrIleHisAlaProAsnLeuSerIleGlnLysIlePhe 493
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2511 ATAGCAAGTGC.....AAACGCTAATGTAAGT..... 2539
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
494  LeuSerAsnSerGlyAspGluAsnPheTyrGluAsnValGluLeuLeuSe 510
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2540 ...ACCATCTCGTAATGATAAC...AGCACCGGTTTAACCATTTCCG 2583
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
510  rLysGlu.....GlnAsnAsnIleProLeuLeuThrLeuSerLysG 524
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2584 AAAAGATGTAACGGTAACAACTAAGCTT.....ACCTCCCA 2621
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
524  LuGlnSerHisLeuHisLeuProAspGlyAsnLeuSerSerHisPheGly 540
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2622 AGACAATAAATATCTCGCCGACGAGGAATGTAACAACCAAGAAAGC 2671
```

seq documentation block:

Sequence 7, Application US/08302832
Patent No. 5603938
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,832
FILING DATE: 16-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US pct/us93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-404
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4287 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-832-7

alignment_scores:

Quality:	144.00	Length:	788
Ratio:	0.364	Gaps:	38
Percent Similarity:	50.254	Percent Identity:	19.670

alignment_block:

US-09-677-752-4 x US-08-302-832-7

Align seg 1/1 to: US-08-302-832-7 from: 1 to: 4287

72 GlyAsnSerTyrCysTirPheValSerLysLeuHisIleThrAspProLys 88
|||||
1208 GGAGCGACCTAC...TGGACGTAAACCTTAAATGTTACTCG... 1249
88 sGluAlaLeuPheLysGlyAspLeuSerIleGlnAsnPheArgp 105
|||||
1250GGTAGTAATTAACCTCTCCATTGACAGC..... 1279
105 heLeuSerPheThrAspCysSerLysGlySerProSerIleLle 121
|||||
1280ACAGGAAGTGCCTCAGCAGTCCAGCAATA... 1309
122 HisGlnLysAsnGlyGlnLeuSerLeuArgAsnGlySerMetSerPh 138
|||||

1310CGCAATGCGAATTAAATGGCATTAATTAAAGCCACTT 1353
138 eCysArgAsnHisAlaGluGlySerGlyLys.....AlaIleSerAla 153
|||||
1354 T.....AAATCGCACAGGCTCAGACGTAACCTTTAGCATCAGACAT 1397
153 sPalaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPheGluAsn 169
|||
1398 CAATAATGCCCTTTAAGAGTAACCGTAACGTCATATTATTAAAGAGAT 1447
170 SerSerLysGlyAsnGlyAlaIleGlnAlaGlnThrPheSerLeu 186
|||||
1448 ATTCAGTCTCAGGGGGGGGAGCGTTAAATTCACACTTAACCGCTCATC 1497
186 ArgAsnValSerPro.....IleSerPheAlaArgAsnArgLysP 201
|||||
1498 TAGCAACATCAAAACCCCTGGCGTAATTATAAATTCGAAACCTTTATG 1547
201 euAsnGlyGlyAlaIleCysSerAsnLeuIleCysSerGlyAsnVal 217
|||||
1548 TCTCAGGAGGCTCAACTTTA.....AATCTCAAGGCTGAAGGTTCAACA 1591
218 AsnProLeuPhe.....PheThrGlyAs 225
|||
1592 GAAACCGGCTTTTCATAGAAATGATTAACTTAACGCCACCGGTTGG 1641
225 nSerAlaThr..... 228
1442 CAATATACCAATCAGACAGTGCAGGGTACCGANTTCACGCTCAACAAG 1691
229AsnGlyGlyAlaIleCysCys 235
|||||
1692 GTGTCGCGACCCAAAACATTAACCTTTAAAGGCGGTAATATCACCTTC 1741
236 IleSerAspLeuAsnThrSerGlu.....LysGlySerLeuSerLeuAlaC 251
|||||
1742 GGCTCTCAAAAAGCCACACAGAAATCAAAAGCAATGTTACATCAATAA 1791
251 sAsn.....GlnGluThrLeuPheAlaSerAsnSerAlaLysLysGlyG 267
|||||
1792 AAACACTACCTCTCTCTCTGTCGGAATTTGCGCAAAACAAATGCG 1841
267 LysAlaIleTyrAlaLysHisMetValLeuArgTyrAsnGlyProValSer 283
|||||
1842 CTTTAATATATAGCAGCAATGTATT.....AATAATGCAACCTTACC 1885
284 PheIleAsnAsnSerAlaLysIleGlyAlaIleAlaIleGlnSerG 300
|||||
1886 ACTGCGGCTCCATTATCAATATAGCCGGAATCTTACTGTTCAAAAGG 1935
300 YGlySerLeuSerIleLeuAlaGly.....GluGlySerV 312
|||||
1936 CGCTAACCTTCAACCTATACCAATTTACACTTTTAATGTACCGGCTCA 1984
312 AlLeuPheGlnAsnAsnSer..... 318
|||||
1985TTTGACACACATGCGCGCTTCAACATTTCCATTCAGACGAGG 2029
319GlnArgThrSerAspGlnGlyLeuValArgAs 329
2030 GCTAATTTAAGATATCATCAATACACCGTAGCTTAATATTACACCAA 2079
329 nAlaIleTyrLeuGluLysAspAlaIleLeuSerSerLeuGluAlaArg 346
|||||
2080 CTCTGATACCACTTACCGCACCATTTATTAAGGCAATATATCAACAAT 2129
346 snGlyAspIleLeuPhePheAsp.....ProIle 355
|||||
2130 CAGCGATTTGATATATTTGATATAAAAGAGACCGCTGAATCCAAATT 2179
356 ValGlnGluSerSerLysGlySerProLeuPro...SerSerLeuG 371
|||
2180 GCGGCAATATCTCACAAAAGGCAATCTCACAATTTCTTGATATA 2229

```
662 LysSerSerAspSerPheIleThrSerThrCluThrThrSerTyrIleAl 678
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2023 ACCGCGGAGTAATTAATTAACCTTCATTCAGACACAGGAGGCGTC 2072
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
678 aThrValGlnAlaGlnLeuAlaThrSerLeuMetLysIleSerAlaGlnA 695
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2073 AACCA..... 2076
695 lacYtyrAsnGluSerIleHisGluLeuLysThrLysTyrArgSerPhe 711
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2077 .....GGTCAGACATACCGAATGAGAAATTAATGGCATACATT 2118
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
712 SerLysCluGlyPheGlySerThrPheSerValAlaValSerGlyGluVa 728
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2119 AATAAGGACACTTAAATATATGACACAGGCTACAGCACTAATTAGCAT 2168
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
728 lCysAlaSerIle...ProIleValSerAsnGlySerGlyLeuPheSer 744
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2169 CAGCGATCAATTAATATGCCCTTAAAGAGTAAC.....GCTA 2203
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
744 ePheSerIlePheSerLysLeuGlnGlyPheSerGlyThrGlnAspGly 760
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2204 ACTAGCATTAATTAATGAGATATTCTAGTCTCAGG..... 2241
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
761 pheCluGluSerSerGlyCluIleArgSerPheSerAlaSerSerPheAr 777
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2242 .....GGGGGTAGCCTTAATTCAACTAAGCCTCATCTAG 2279
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
777 gAsnIleSerLeuProIleGlyIleThrPheGluLysLysSerGlnLysT 794
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2280 CACATACAAACCCCT...GGCGTAATTATA.....AATCTCAAAAC 2319
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
794 hArgThrTyrTyrPheLeuGlyAlaTyrIleGlnAspLeuLysArg 810
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2320 .....TTAATGTCAGAGGGGTCACTTAATCTCAAGGCT 2358
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
811 AspValGluSerGlyProValIleLeuLeuLysAsnAlaValSerTyrAs 827
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2359 GAAAGGTTCAAGAAACCCCTTTTCATAGAAATGATTAACCTAATA 2408
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
827 pAlaProMetAlaAsnLeuAspSerArg 836
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2409 CCCCAACCGGTGCAATATTAACAATCAGA 2436
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: /cgn2_6/prodata/2/lna/5A_COMB.seq:US-08-038-682-7
seq_documentation_block:
: Sequence 7, Application US/08038682
: Patent NO. 5549897
: GENERAL INFORMATION:
: APPLICANT: BARENKAMP, STEPHEN J
: APPLICANT: ST. GEORGE III, JOSEPH W
: TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Maltare, Ltd
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/038,682
: FILING DATE: 16-MAR-1993
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
```

```
NAME: BERKSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-293
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4287 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-038-682-7

alignment_scores:
Quality: 144.00 Length: 788
Ratio: 0.364 Gaps: 38
Percent Similarity: 50.254 Percent Identity: 19.670

alignment_block:
US-09-677-752-4 x US-08-038-682-7 ..
Align seg 1/1 to: US-08-038-682-7 from: 1 to: 4287

72 GlysAsnSerTyrCysTrpPheValSerLysLeuHisIleThrAspProly 88
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1208 GAGCGACCTAC...TGAAGCTAACCACTTAATGTACTCTG.... 1249
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88 sGluAlaLeuPheLysGlyLysGlyAspLeuSerIleGlnAsnPheArgp 105
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1250 .....GGTAGTAATTAATTAACCTCCCTACACAC..... 1279
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
105 heLeuSerPheThrAspCysSerLysGlyLeuSerSerProSerIleIle 121
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1280 .....ACAGCAAGTCGCTCAACAGCTCCACAGCTA... 1309
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122 HisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPh 138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1310 .....CGCATGCAAGATTAATGCGCATACATTATTAAGCCACTTT 1353
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
138 eCysArgAsnHisAlaGluGlySerGlyGly....AlaIleSerAlaA 153
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1354 T.....AATATCGCAACAGGCTCAACAGCTACTTAGCATCAAGGAT 1397
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1398 CAATAATGCCCTTAAGAGTAGTACCTACTACGCAATTATTAATGAACAT 1447
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
170 SerSerLysGlyAsnGlyGlyAlaIleGlnAlaGlnThrPheSerLeuSe 186
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1448 ATTTAGTCTCAGGGGGGTAGCGTTAATTTCAACTTAACGCTCATC 1497
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
186 fArgAsnValSerPro.....IleSerPheAlaArgAsnArgAlaAspL 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1498 TAGCAACATACAAACCCCTGGCGTAAATTAATAATCTCAAACTTAATG 1547
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 euAsnGlyAlaAlaIleCysCysSerAsnLeuIleCysSerGlyAsnVal 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1548 TCTCAGAGAGGTCAACTTTA.....AATCTCAAGCTGAAGCTTCACA 1591
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218 AsnProLeuPhe.....PheThrGlyAs 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1592 GAAACCGCTTTTCAATAGAAATGATTTAACTTAAGCCACCGGTGG 1641
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
225 nSerAlaThr..... 228
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1642 CAATATTAACAATCAGCAAGTGCAGGTAACGATTCAGCGCTCAACAAG 1691
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
229 .....AsnGlyGlyAlaIleCysCys 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1692 GTTCGCAAGCCAAAAAACAATAACTTTAAAGGGGTAATATCTACCTTC 1741
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

486	AsnLeuSerIleGlnIlyIlePheLeuSerIxnSerGlyAspGluAsnPr	502
2484		2533
502	etYcIu.....AsnValGluLeuLeuSerIys.....	511
2534	AGCAGCCGACGCGGGAATTGCTAAATGTGTGCGCCACGGACGTAAACA	2583
512GluGlnAsnAsnIleProLeu.....	518
2584	TTACAGAGAGAGGAAAGAAAGATTTCAGCGGTCAACACGATCTTTAAACGGA	2633
519LeuThLeuSerIysGluInSerHisLeuHisLeuPr	531
2634	ACCGGTAAAGCTGTGTAATATCATTTTCATCACTGTAATATATTA.....	2675
531	oAspGlyAsnLeuSerSerHisIlePheGlyTyrGlnGlyAspTrpThrPheS	548
2676		2724
548	et.....TrpIlyAspSerAspGluGly	555
2723	ACCAAACTACGAGAAAGAACACTGCTATTGTGGCAAAACCGACCATGATTCG	2774
556	HisSerLeuIle.....AlaAsnTrpThrPr	564
2775	CACGTGACGCTACGTCGTCCTTAATCTAACAGAACGCCAAATTTTACCTT	2824
564	oLysAsnTyrValProHisProGluArgGlnSerThr.....	576
2825	TATTAAATGATTCATTCCACCAATAGCCAAAGCGCTTAACAACACGATATGAA	2874
576	576
2875	GCTGTGACGGGTGAATTTTAACGGCGTAATGGCACATGTCATTCAT	2924
577LeuValAlaAsnThrLeuTr	583
2925	CTCAAAAGAGAGCGCAAGATTAAATTTCAAATTTAAACCAACGACACAT	2974
583	PasnThrTyrSerArgPheMetGlnAlaValGlnSerMetLeaSnThrIleA	600
2975	GAACAA.....ACCAAACTTTTACCAATTCGCGTTTATACCAATATTCACAG	3021
600	IanHisGlyGlyLeuTyrLeuThrPheGlyTrpGlySerAlaValSerAsn	616
3022	CCACTGGTGG.....GCTCGTTTTTTTTTTCAT	3050
617	LeuPheTyrAlaHisAspSerSerGlyLysProIleAsp.....AsnTrpHi	632
3051	AAATATGCCAACCATTTGCGAGAGGGCGTCACTAAATAAGAGCAAT	3100
632	HisAspSerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspA	649
3101	TATATCTCTAACGGCGGTAATTTTACCTTAATTCGATGTGGCGGCG	3150
649	sPHisSerPheCysLeuAlaIaGlyGlnLeuLeuGlyLysSerSerAsp	665
3151	ATGACGCTTTTAAATC.....AACAAAGC	3176
666	SerPheIleThrSerThrGluThrThrSerTyrIleAlaThrValGlnAl	682
3177	TTTACCATTAATGCAACCAATTCA.....	3200
682	acGlnLeuAlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrAsnG	699
3201AAT	3204
699	IuSerIleHisGluLeuLysThrIys.....TyrArgSerPheSerIysGlu	714
3205	TGAGCTTCAGACAGACAAAGATGATTTTATGACGGGTACGACAGCAAT	3254


```

1656 GATTCTGAAGCAAGATAAAGCGCATATTGTTCTTCCGCCAAGA 1705
273 smevValleuArgTyrasnGlyProValSerPheIleAsnSerIal 290
1706 GCGTGAAGCGGAAATTGGCGGTAAATTTCGCTCAAAATCAGCACTA 1755
290 ysiIleGlyValIleAlaIleGlnSer..... 299
1756 AA...GGCGGCAAGCTGATGATAAGTCCGATAAGTCACATTAAACA 1802
300 GlyIleSerSerIleLeuAlaGlyGlySerValLeuPheGlnAs 316
1803 GGTCCACTTATTCGACCTTTCAGTAAGAAGGCGCAAACTTACCTCG 1852
316 naenSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrL 333
1853 CGGTGACGAGCGC.....GGCGAAGTAAACGGCGATTCAAT 1890
333 euGluIysAspAlaIleLeuSerSerLeuGluAlaArgAsn..... 346
1891 TAGCAAGAAACCTCTTAGAAAAGGCTCAACCATCAATGTTCAGGC 1940
347 .....GlyAspIleLeuPheAs 353
1941 AAAGAAAAGCGGACCGCCATTTGTGTGGCGCATATTGGCTTAATTGA 1990
353 proIleValGlnIleUserSerSerLysGlyUserProLeuProSerSerL 370
1991 C.....GGCAATA 1998
370 euGlnAlaSerValThrSerProThrProAlaThrAlaSerProLeuVal 386
1999 TTACGGCTCAAGTAGTGTGATATCGCTAAACCGGTGGT.....TTT 2042
387 ILeGlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSe 403
2043 CTGGAGACATCGGGCATTTATTCATTCAGACCAATGCAATGTGTAA 2092
403 rGluGluGlu.....LysThrProAspAsnLeuThrSerGlnLeuGlnG 418
2093 AACAAAAGAGGTGGTCTAGACCCGTGATGATGATGACAAATGACCGCAG 2142
418 InProIleGluLeuLysSerGlyArgLeuValIleuLysAspArg..... 432
2143 ACCCCCTTCGCATATATACCGGT.....ATAATGATGAATTTCCCA 2183
433 .....AlaValIleUserAlaProSerLeuSerGlnAspProGlnAl 446
2184 ACAGGCACCGGTGACCAACGACCCCTAAAAAAATACGCAATCAAAAC 2233
446 aleuLeuIle..... 449
2234 AACGCTAACCAATACACTATTTCAAATTATCTGAAAAACGCGCTGACAA 2283
450 ....MecGluIaGlyThrSerLeuLysThrSerSerAspLeuLysLeu 464
2284 TGAATATATACGGCATCAAGAAACTTACCCTTAATAGCTCAATCAACATC 2333
465 AlaThrLeuSer...IleProLeuHisSer..... 473
2334 GGAAGCAACTCCCACTTAATCTCATAGTAAGTCACACCTGCGCGAGG 2383
473 ..... 473
2384 CCTTCAGATTGATGAGATATTACTTTAAAGCGGAAATTTAACCATTT 2433
474 .....LeuAspThrGluLysSerValThrIleHisAlaPro 485
2434 ATTTCGGCGGATGGGTGATGTTCATAAAATATTACGCTTATCAGAGGT 2483
486 AsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspGluAsnPh 502
2484 TTTTAAATATTAACGCCCTTCGTAAGCTTTTGAAGGTGCAAAATAACA 2533

```

```

502 eTyrGlu.....AsnValGluLeuLeuSerLys..... 511
2534 AGCAGCGGACGGCGCAATCTTAATTTGTCGCCACGGGACGTATACCA 2583
512 .....GluGlnAsnAsnIleProLeu..... 518
2584 TTACAGAGAGAGGAAAAAGATTTCAGCGCTAACCAAGTATCTTAAACGCA 2633
519 .....LeuThrLeuSerLysGluGlnSerHisLeuHisLeuPr 531
2634 ACGGTAAAGCTCTGAATATCATTCATTCATCAGTAGAATATTTA..... 2675
531 oAspGlyAsnLeuSerSerHisPheGlyTyrGlnGlyAspThrPheS 548
2676 .ACCCACAAATCTTAGTGCCAAATTAACATATCTGGAAATATACATTA 2724
548 er.....TyrLysAspSerAspGluGly 555
2725 ACCAACTACAGAGAAAGACACCTCGTATTTGGCAACACCATGATTCG 2774
556 HisSerLeuIle.....AlaAsnThrPhePr 564
2775 CACTGGAACGTCACTGCTTAATCTAGAGACAGCGCAATTTTACCTT 2824
564 oLysAsnTyrValProHisProGluArgGlnSerThr..... 576
2825 TATTAAATACATTTCAGCAATAGCAAGGCTTAACACACAGATATAGAA 2874
576 ..... 576
2875 GCTCTGACGGGTGAATTTTAAACGGCTAAATGGCAACATGTTCATCAAT 2924
577 .....LeuValAlaAsnThrLeuTr 583
2925 CTCAAAGAGAGCGAAAGTTAATTTCAATTAACCAACGAGAACAT 2974
583 PAsnThrTyrSerAspMetGlnAlaValGlnSerMetIleAsnThrIleA 600
2975 GAAACAA...AGCAACCTTACCAATTCGGTTTTCAGCAATATCAAG 3021
600 IAsnIleGlyAlaTyrLeuPheGlyThrTrpGlySerAlaValSerAsn 616
3022 CCACGTGGGCG.....GGCTCGTTTTTTTCAT 3050
617 LeuPheTyrAlaHisAspSerSerGlyLysProIleAsp...AsnThrPhI 632
3051 ATATATGCCAACCATTTCTGGCAGAGGGCTGAGTTAAATGAGTGAAT 3100
632 SHisArgSerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspA 649
3101 TAAATCTCTAACGGCGTAAATTTTACCTTAATTCATGTTCGGCGCG 3150
649 sPHisSerPheCysLeuAlaIaGlyGlnLeuLeuGlyLysSerSerAsp 665
3151 ATGACGCTTTTAAATC.....AACAAAGAC 3176
666 SerPheIleThrSerThrGluThrThrSerTyrIleAlaThrValGlnAl 682
3177 TTAACCAATAATGCACCAATTTCA..... 3200
682 aGlnLeuAlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrAsnG 699
3201 .....AAT 3204
699 IuSerIleHisGluLeuLysThrLys...TyrArgSerPheSerLysGlu 714
3205 TCAGCCTCAGACAGAGAAAGATGATTTTATGACGCGTACGACGCAAT 3254
715 GlyPheGlySerThrHisSerValAlaValSerGly 726
3255 GGCATCAATTCACCTCAACATATCATTCCTGGGC 3290

```



```

1415 :|||||:.....: |||||:
      CGCAGTGAATCTATT.....GTGCAAG 1443
150  leSerAlaSpAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
      :....:
1444  TGAAG..... 1448
167  GluGluAsnSerSerLySGlyAsnGlyAlaIleGln.....Al 180
      :|||||:.....: |||||||:.....:
1449  AACGAGCGCTGATTAGCGTAATGCTGCGACATTTCTTACTTCGCCAG 1498
180  aGlnThrPheSerLeuSerArgAsnValSerPro.....I 192
      :|||:.....: |||:.....:
1499  GCAGAAATGCACCATCAGCATATATTAACCAACCATTTACTTACAGCA 1548
192  leSerPheAlaArgAsnArgAlaAspLeuAsnGlyAlaIleCysGly 208
      :|||||:.....: |||:.....:
1549  TTGCGCGCCCTGAAATGAGACG...GTCAATCTGGCGGATATTTTCC 1595
209  SerAsnLeuIleCysSerGlyAsnValAsnProLeuPheThrGlyAs 225
      :|||:.....: |||:.....:
1596  AAA.....GCCGCTACATTAATGTC.....CG 1618
225  nSerAlaThrAsnGlyAlaIleCysCysIleSerAspLeuAsnThrS 242
      :|||||:.....:
1619  TGCTGCCACT..... 1628
242  erGluLySGlySerLeuSerLeuAlaCysAsnGlnIleuThrleuPheAla 258
      :|||||:.....: |||:.....:
1629  ....ATTCGAACCAAGTAACTTCTGCT 1655
259  SerAsnSerAlaLySGlyLySGlyAlaIleTyr.....AlaLySIII 273
      :|||||:.....: |||:.....:
1656  CATTCGTAGCAAGATAAAGCGGCAATATTTCTTCCGCCAACA 1705
273  sweValLeuArgTyrAsnGlyProValSerPheIleAsnSerAlaL 290
      :|||:.....: |||:.....:
1706  GCGTGAACGGAAATTTGGCGGTATTTCCGCTCAAAATCAGCAAGCTA 1755
290  ysIIeGlyAlaIleAlaIleGlnSer..... 299
      :|||:.....: |||:.....:
1756  AA...GGCGGCAAGCTGATGATNAAGTCGATNAAGTCACATTAAACA 1802
300  GlyGlySerLeuSerIleLeuAlaGlyGlnGlySerValLeuPheGln 316
      :|||||:.....: |||:.....:
1803  GGTGAGATTATGACCTTTCAAGTAAGAAGGCGAAGCAACTTACTTCG 1852
316  nasnSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrL 333
      :|||||:.....: |||:.....:
1853  CGGTGACGAGCGC.....GCCGAAGGTAAACCGCATTCAT 1890
333  euGluLyAspAlaIleLeuSerSerLeuGluAlaArgAsn..... 346
      :|||:.....: |||:.....:
1891  TAGCAAGAAACCTTTAGAAAAAGCTCAACCATCATGTATTCAGGC 1940
347  ....GlyAspIleLeuPhePheAs 353
      :|||||:.....: |||:.....:
1941  AAGAAAAAGCGGACGCCCTATTGCTGGCGCATTTGCTTAATTGA 1990
353  pProIleValGlnGluSerSerLySGluSerProLeuProSerSerL 370
      :|||:.....: |||:.....:
1991  C.....GGCAATA 1998
370  euGlnAlaSerValThrSerProThrProAlaThrAlaSerProLeuVal 386
      :|||||:.....: |||:.....:
1999  TTAACGCTCAAGTAGTGATATCGCTAAACCGGTGT.....TTT 2042
387  IleGlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSe 403
      :|||||:.....: |||:.....:
2043  GTGAGACATCGGGCATTTATTCATTCATGACAGCAATGCAATGTATA 2092
403  rGluGluGlu.....LysThrProAspAsnLeuThrSerGlnLeuGlnG 418
      :|||||:.....: |||:.....:
2093  AACAAAAAGTGTGTGCTAGACCTGATGATGATNAACATTGAAGCCGAG 2142
418  InProIleGluLeuLySGlyArgLeuValLeuLyAspArg..... 432
      :|||||:.....: |||:.....:
2143  ACCCCTTGCAATATATACCGT.....ATTAATGATGAATTCCTCA 2183
433  ....AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAl 446
      :|||:.....: |||:.....:
2184  ACAGGACCGGTGAAGCAAGCAGACCTTAACCAAAAAATACGAACTMAAC 2233
446  aLeuLeuIle..... 449
      :|||:.....: |||:.....:
2234  AACGCTAACCAATACAACTATTTCAAATTATCTGAAAAACCGCTGACAA 2283
450  ....MetGluAlaGlyThrSerLeuLyThrSerSerAspLeuLyLeu 464
      :|||||:.....: |||:.....:
2284  TGAATATACCGCATCAAGAAACTTACCGTTAATAGCTCAATCAACATC 2333
465  AlaThrLeuSer...IleProLeuHisSer..... 473
      :|||:.....: |||:.....:
2334  GGAACCACTCCCACTTAATTCTCATAGTAAAGCTCAGCGTGGCGGAG 2383
473  .... 473
2384  GGTTCAGATTGAGATATTACTTCTAAAGCGGAAATTTAACCATTT 2433
474  ....LeuAspThrGluLySerValThrIleHisAlaPro 485
2434  ATTCGTGGCGGATGGTGTGATGTTCTATAAAATTTACCTTGATCAGGCT 2483
486  AsnLeuSerIleGlnLyIlePheLeuSerAsnSerGlyAspGluAsnPh 502
      :|||||:.....: |||:.....:
2484  TTTTAAATATTACCGCGCGCTTCCTTGTGAGAGTGCAGAAATACAA 2533
502  eTyrGlu.....AsnValGluLeuLeuSerLyS..... 511
      :|||:.....: |||:.....:
2534  AGCAGCGACGCGCAAAATGCTAAATTTGCGCCAGCGCACTGAACA 2583
512  ....GluGlnAsnAsnIleProLeu..... 518
2584  TTACAGCAGAGGAAAAACATTTCAAGGCTACACAGATATCTTTAAACGA 2633
519  ....LeuThrLeuSerLySGlnGlnSerHisLeuHisLeuPr 531
      :|||||:.....: |||:.....:
2634  ACGGTAAGGCTGATATATTCATTTCAATCAGATATATTTA..... 2675
531  aspGlyAsnLeuSerSerHisPheGlyTyrGlnGlyAspThrPhePheS 548
      :|||||:.....: |||:.....:
2676  ACCCACAATCTTAGTGACAAATTAACATATCTGGATATTAACAATTA 2724
548  er.....TrpLyAspSerAspGlnGly 555
      :|||:.....: |||:.....:
2725  ACCAACTACGAGAAAGAACACCTCTGATTGGCAACCGACCATGATTG 2774
556  HisSerLeuIle.....AlaAsnTrpThrPr 564
      :|||:.....: |||:.....:
2775  CACTGGAAGCTGAGTCGTATATCTAGAGACAGCGCAATTTTACTT 2824
564  OLyAsnTyrValProHisProGluArgGlnSerThr..... 576
      :|||||:.....: |||:.....:
2825  TATTAATATACATTTCAGCAATAGCAAGGCTTAACAACACAGATAGAA 2874
576  .... 576
2875  GCTTGACAGGGTGAAATTTTAAGGCGTAAATGCAACATGCTCATTCAT 2924
577  ....LeuValAlaAsnThrLeuTr 583
2925  CTCGAAGAAGGACGCAAAATTAATTTCAATTAACCAACGACAGACAT 2974
583  pasnThrTyrSerAspMetGlnAlaValGlnSerMetIleAsnThrIleA 600
      :|||||:.....: |||:.....:
2975  GAACACA...AGCAAACTTTACCAATTCGGTTTTCACCAATATACAG 3021

```

```
REFERENCE/DOCKET NUMBER: 1038-404
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 9323 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-832-6

alignment_scores:
Quality: 148.50      Length: 812
Ratio: 0.408         Gaps: 36
Percent Similarity: 44.828      Percent Identity: 18.227

alignment_block:
US-09-677-752-4 x US-08-302-832-6

Align seg 1/1 to: US-08-302-832-6 from: 1 to: 9323

73 AsnSerTyrCysTrpPheValSerLysLeuHisIleThrAspProLysG1 89
1218 AACGGACAGCTCTTTTATTCACCAACCAATGGATTCACATAGCTAAGA 1267
89 uAlaLeuPheLysGluLysGly.....AspLeuSerI 100
1268 CCGCATTTATTAACACTAATAGCTTACGGCTTACGCTAGACATTTCTA 1317
100 LeG1AsnPhaArgPheLeuSerPheThrAspCysSerLysG1user 116
1318 ACGAAACATCAAGCGCGCTAATTTACCTTCAGACCAACCAAGATAA 1367
117 SerProSerIleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsn 133
1368 CCGCTCGCTGAATTT...GTGAATCAGCGTTAATTACTGTCGTAAGA 1414
133 nGlySerMetSerPheCysArgAsnHisAlaGluGlySerGlyAlaI 150
1415 CGCGAGTGTAAATCTTAT.....GGTGGCAAG 1443
150 LeSerAlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaP 166
1444 TCAAA..... 1448
167 GluGluAsnSerSerLysGlyAsnGlyAlaIleGln.....Al 180
1449 AACGAGCGTGTGATTAGCTTAATGTGGCAGCATTTCTTACTCGCAGC 1498
180 aGlnThrPheSerLeuSerArgAsnValSerPro.....I 192
1499 GCATAAAATCACCATCAGCATATATAACCCACCATTTACTTACAGA 1548
192 LeSerPheAlaArgAsnArgAlaAspLeuAsnGlyAlaIleLecysG 208
1549 TTGCGCGCGCTTAATAATGAGCG...GTCAATCTGGCGCATATTTTTC 1595
209 SerAsnLeuIleCysSerGlyAsnValAsnProLeuPhePheThrGly 225
1596 AAA.....GGCGCTACATTAATCTC.....CG 1618
225 nSerAlaThrAsnGlyAlaIleLecysCysIleSerAspLeuAsnThr 242
1619 TGCTGCGCT..... 1628
242 eGCluLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAl 258
1629 .....ATTTCAGAACCAAGTAACCTTTCTCT 1655
259 SerAsnSerAlaLysGluLysGlyAlaIleTyr.....AlaLysHis 273
```

```
1656 GATTCTGTAGCAAGAAATTAACGGCATATTTGTTCCGCCCAAGA 1705
273 sMetValLeuArgTyrAsnGlyProValSerPheIleAsnSerAla 290
1706 GGGTGAAGCGGAATGTGGCGGTATTTCCGCTCAAAATCAGACAGTA 1755
290 ySileGlyAlaIleAlaIleGlnSer..... 299
1756 AA...GGCGCAGCTGATGATTAAGCTCCATTAACATCATTAAGA 1802
300 GlyGlySerLeuSerIleLeuAlaGlyGluGlySerValLeuPheGln 316
1803 GGTGCAATATGACCTTACAGTAAAGAGGGGGAACCTTACCTTGG 1852
316 nAsnSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyr 333
1853 CGGTGACGAGCGC.....GGCAAGGTAAACCGCATTCAT 1890
333 eGluLysAspAlaIleLeuSerSerLeuGluAlaArgAsn..... 346
1891 TAGCAAGAAACCTCTTGAAGAAAGCGTCAACATCATATGATCAGC 1940
347 .....GlyAspIleLeuPheAsn 353
1941 AAGAAAAAGCGGACGCCCTATTGTGGGGGATATGCGTTAATTA 1990
353 pProIleValGlnGluSerSerLysGluSerProLeuProSerSerL 370
1991 C.....GGCAATA 1998
370 eGluAlaSerValThrSerProThrProAlaThrAlaSerProLeuVal 386
1999 TTAACCTTCAGCTAGTGTGATATCGTAAACCGGTGT.....T 2042
387 IleGlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLeu 403
2043 GTGGAGACATCGGGCATTTATTTATTCATTGACAGCAATGATGTAA 2092
403 rGluGluGlu.....LysThrProAspAsnLeuThrSerGlnLeuGln 418
2093 AACAAAGAGGTGGTGTGATGACCTGATGATGATCAATGAGCGGAAG 2142
418 InProIleGluLeuLysSerGlyArgLeuValLeuLysAspArg..... 432
2143 ACCCCCTTCGCAATATACCGT.....ATAATGATGAATTCACA 2183
433 .....AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAl 446
2184 ACGGACCGGTGAAGCAAGCAACCTTAATAAAATAGCAACTCAAAAC 2233
446 aLeuLeuIle..... 449
2234 AACGCTAACCATACACATTTCAATTAATCTGAAAAAGCGCTGACAA 2283
450 ....MetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeu 464
2284 TGAATATACCGCATCAAGAAACCTTACCGTTAATAGCTCAATCAATC 2333
465 AlaThrLeuSer...IleProLeuHisSer..... 473
2334 GGAACCACTCCACTTAATTTCCATAGTAAGGTACAGCGTGGGAGG 2383
473 ..... 473
2384 CATTGAGATTGATGAGATATTACTTAAGCGCGGAATTTAACCATTT 2433
474 .....LeuAspThrGluLysSerValThrIleHisAlaPro 485
2434 ATTCTGGCGGATGGTGTGATTTCAATAAAATATATACCTGATCAGCGT 2483
486 AsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspGluAsnPh 502
|||||.....
```

```

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-038-682-6
seq_documentation_block:
  Sequence 6, Application US/08038682
  Patent No. 5549897
  GENERAL INFORMATION:
  APPLICANT: BARENKAMP, STEPHEN J
  APPLICANT: ST. GEME III, JOSEPH W
  TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEININ
  TITLE OF INVENTION: OF NON-TYREABLE HAEMOPHILUS
  NUMBER OF SEQUENCES: 8
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: Shoemaker and Mattare, Ltd
  STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
  STREET: Bldg. 1
  CITY: Arlington
  STATE: Virginia
  COUNTRY: U.S.A.
  ZIP: 22202-0286
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  OPERATING SYSTEM: IBM PC compatible
  SOFTWARE: Patent Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/038,682
  FILING DATE: 16-MAR-1993
  CLASSIFICATION: 424
  ATTORNEY/AGENT INFORMATION:
  NAME: BERKSTRESSER, JERRY W
  REGISTRATION NUMBER: 22,651
  REFERENCE/DOCKET NUMBER: 1038-293
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: (703) 415-0810
  TELEFAX: (703) 415-0813
  INFORMATION FOR SEQ ID NO: 6:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 9323 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: DNA (genomic)
  US-08-038-682-6

alignment_scores:
  Quality: 148.50      Length: 812
  Ratio: 0.408         Gaps: 36
  Percent Similarity: 44.828   Percent Identity: 18.227

alignment_block:
  US-09-677-752-4 x US-08-038-682-6
  Align seg 1/1 to: US-08-038-682-6 from: 1 to: 9323

73 AsnserTTCyCstGrpHeValSerLysLeuHisILethrAspProLysGI 89
|||||
|||||
1218 AACGACACAGCTCTTTTAATCAACCCCAAAATGTAATCACAATAGCTAAGA 1267
|||||
|||||
89 uAlaLeuPheLysGILuLysGly .....AspLeuSerI 100
|||||
|||||
1268 CGCAATTATTAACTAATAGGCTTTACGGCTTCTACCGTAGACATTCTTA 1317
|||||
|||||
100 leGlnAsPheArGrPheLeuSerPheThrAspCysSerSerLysGIuSer 116
|||||
|||||
1318 ACGAAACACATCAAGCGCCGTAATTTTCACCTTCGAGCAACCAACGATGAA 1367
|||||
|||||
117 SerProSerIleIleHisGlnLysAsnGlyVbIleuSerLeuAtGnaSns 133

```

```

577 .....LeuValAlaasnThrLeuTrpAs 584
2488 AAGAAGGAGCGAAGTAAATTCAATTAAACCAAGACGACACTGA 2537
584 nThrTyrSerAspMetGlnAlaValGlnSerMetIleasnThrIleAla 601
2538 CACA...ACCAAACTTTACCAATTCGGCTTTTACGCATATACACGCCA 2584
601 IscGlyValTyrLeuPheGlyThrTrpGlySerAlaValSerAsnLeu 617
2585 CTGGTCG.....GGCTCTGTTTTTTTTCATATA 2613
618 PheTyrAlaHisAspSerSerGlyLysProIleAsp...AsnThrPheIst 633
2614 TTGGCAACCATCTGGCAGAGGGCTGATTAAATAGTGAATTAA 2663
633 sarGSerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAsp 650
2664 TATCTCTAACGGCGGCGTAATTTTACCTTAATTCCTATGTTCCCGCAG 2713
650 IsSerPheCysLeuAlaAlaGlyGlnLeuLeuGlyLysSerSerAspSer 666
2714 ACCCTTTTAAATTC.....AACAAAGACCTTA 2739
667 PheIleThrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaG 683
2740 ACCATAATGCACCAATCA..... 2760
683 nLeuAlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrAsnGlu 700
2761 .....AATTTTCA 2767
700 eTlleHisGluLeuLysThrLys...TyrArgSerPheSerLysGluGly 715
2768 GCGTAGACAGCAGCAAGATGATTTTGTAGCGGTACCCAGCAATGCC 2817
716 PheGlySerThrPheSerValAlaValSerGly 726
2818 ATCATTCACCTACACATATCCATCTCGGCC 2850

seq_name: /cgn2_6/prodata/2/ina/5B_COMB.seq:US-08-617-697-3
seq_documentation_block:
: Sequence 3, Application US/08617697
: Patent No. 5977336
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Matzare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/617,697
: FILING DATE: 01-APR-1996
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 05-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166

```

```

FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
: NAME: Berkstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-557
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4937 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-617-697-3

alignment_scores:
: Quality: 148.50 Length: 811
: Ratio: 0.408 Gaps: 35
: Percent similarity: 44.883 Percent identity: 17.879

alignment_block:
US-09-677-752-4 x US-08-617-697-3
Align seg 1/1 to: US-08-617-697-3 From: 1 to: 4937

73 AsnSerTyrCysTrpPheValSerLysLeuHisIleThrAspProLysG 89
||||| :|||||
778 AACGACACAAGCTCTTTTATATCAACCCAAATGCTATCACATAGCTAA 827
89 uAlaLeuPheLysGluLysGly.....AspLeuSerI 100
||||| :|||||
828 CCGCATTTATTAACATAATAGCGCTTACGCGCTTACGCTAGCAATTGTA 877
100 LeuGlnAspPheArgPheLeuSerPheThrAspCysSerSerLysGluSer 116
||||| :|||||
878 ACGAAACATCAAGCGCGTAAATTTCCACCTTCGACCAACCAAAATATAA 927
117 SerProSerIleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAs 133
||| :|||
928 GCGCTCGCTGAATTT...GTGAATCAGCGTTTAATTACTGTCGGAAGA 974
133 nGlySerMetSerPheCysArgAsnHisAlaGluLysGlyGlyAlaI 150
||||| :|||||
975 CGGCAGTGTAATCTTATT.....GGTGCAGAG 1003
150 IsSerAlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
||||| :|||||
1004 TGAAA..... 1008
167 GluGlnAsnSerSerLysGlyAsnGlyAlaIleGln.....Al 180
||||| :|||||
1009 AACGAGCGCTGATGATGACGTAAATGCTGCGACATTTCTTACTCGCAG 1058
180 aGlnThrPheSerLeuSerArgAsnValSerPro.....I 192
||| :|||
1059 GCAAAATAATCACATCACGCGATATATTAACCAACCATTAATTACACA 1108
192 IsSerPheAlaArgAsnArgAlaAspLeuAsnGlyAlaIleCysGly 208
||||| :|||||
1109 TTGCGCGCGCTGAATAATGAGCG...GTCAATCTGGGCGATATTTTGGC 1155
209 SerAsnLeuIleCysSerGlyAsnValAsnProLeuPhePheThrGlyAs 225
||| :|||
1156 AAA.....CGGCTAATATATGTC.....CG 1178
225 nSerAlaThrAsnGlyAlaIleCysCysIleSerAspLeuAsnThr 242
||||| :|||||
1179 TGCTGCCACT..... 1188
242 eArgLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAla 258

```

2047 TTAATATTACCGCCGCTTCCGTACGTTTGAAGTGGAAATACAAAGC 2096
503 rglu.....AsnValGluLeuLeuSerLys..... 511
2097 ACGGACAGCGGGAATGCTAAATTTGTCCGCCAGGACATGTAAACATTAA 2146
512GluGlnAsnAsnIleProLeu..... 518
2147 CAGGAGAGGAAAGATTTCAGGCGCTAACACGCTATCTTAAACGGAAGC 2196
519LeuThrLeuSerLysGluGlnSerHisLeuHisLeuProAs 532
2197 GGTAAAGCTCTGAATATATCATTCATCAGTGAATTA.....AC 2237
532 pClyAsnLeuSerSerHisPheGlyTYrGlnGlyAspTrpTrpPheSer. 548
2238 CCACATCTTACGTGGACATTAACATATCTGGAAATATACAAATTAAAC 2287
549TrpLysAspSerAspGluGlyHis 556
2288 AACTACGAGAAAGAACCTCGTATTTGGCAACACGACATGATTCGCAC 2337
557 SerLeuIle.....AlaAsnTrpThrProly 565
2338 TCGAAGCTCAGTCTCTTAATCTAGACAGCGGCAAAATTTTACCTTTAT 2387
565 sAsnTrpValProHisProGluArgGlnSerThr..... 576
2388 TAAATATACATTTACAGCAATAGCAAAAGGCTTACACACAGTATAGAGCT 2437
576 576
2438 CTGACGGGTGAATTTTAAAGCGGTAAATGSCACATGTCATTCAATCTC 2487
577LeuValAlaAsnThrLeuTrpAs 584
2488 AAGACAGACGCAAGTTAATTCAATTTAAACCAACAGCAGACATGCA 2537
584 nThrTySerAspMetGlnAlaValGlnSerMetLeuAsnThrIleAlaH 601
2538 CACA...AGCAAACTTTACCAATTCGCTTTTACCAATATACACAGCA 2584
601 IsglyGlyAlaValIleuPheGlyThrTPGlySerAlaValSerAsnLeu 617
2585 CTGGTGGG.....GGCTCTGCTTTTGTGATATA 2613
618 PheTYrAlaHisAspSerSerGlyLysProIleAsp..AsnTrpHisHl 633
2614 TATGCCAACCATCTGGCAGAGCGGCTGACITTAATAATCAGTGAATTTAA 2663
633 sArgSerLeuGlyTYrLeuPheGlyIleSerThrHisSerLeuAspAspH 650
2664 TATCTCTAACGGCTAATTTTACCTTAATTCCTATTCGCGGCGATG 2713
650 IserPheCysLeuAlaIaGlyGlnLeuLeuGlyLysSerSerAspSer 666
2714 ACGCTTTTAAATC.....ACCAAGACCTTA 2739
667 PheIleThrSerThrGluThrThrSerTYrIleAlaThrValGlnAlaI 683
2740 ACCATATAAGCAACCAATTCA..... 2760
683 nLeuAlaThrSerLeuMetLysIleSerAlaGlnAlaCysTYrAsnGluS 700
2761AATTGA 2767
700 eTlleHisIleuLeuLysThrLys...TYrArgSerPheSerLysGluGly 715
2768 GCGCTAGACAGACAAAGATGATTTTAAAGCGGTACGACGCAATGCC 2817
716 PheGlySerTrpHisSerValAlaValSerGly 726
2818 ATCAATTCACTACACATATCATTCGCGC 2850

seq_name: /cgn2/6/prodata/2/ina/58_COMB.seq:us-08-728-470-3
seq_documentation_block:
: Sequence 3, Application US/08728470
: Patent No. 5928651
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: of No. 5928651-Typable Haemophilus
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Mattare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/728,470
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Berksstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-633
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0810
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4937 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-728-470-3

alignment_scores:
Quality: 148.50 Length: 811
Ratio: 0.408 Gaps: 35
Percent Similarity: 44.883 Percent Identity: 17.879

alignment_block:
US-09-677-752-4 x US-08-728-470-3 ..
Align seg 1/1 to: US-08-728-470-3 from: 1 to: 4937

73 AsnSerTYrCysTrpPheValSerLysLeuHisIleThrAspProLysG 89
|||||
778 AAGCGACAGAGTCTTTTATATCAACCAATGCTATACATAGGTAAAGA 827
|||||
89 uAlaLeuPheLysGluLysGly.....AspLeuSerI 100
|||||
828 CCAATTTATACACTAATGCGCTTACGGCTCTACGCGACAGCATTTCA 877
|||||
100 IecGlnAsnPheArgPheLeuSerPheThrAspCysSerSerLysGluSer 116

259 SerAnserAlaLysGluLysGlyAlaIleTyr.....AlaLysHI 273
 1216 GATTCTGTAAAGCAAGATAAAGCGCAATATTGTTTCGCGCAAA 1265
 273 sMetValLeuArgTyrAsnGlyProValSerPheIleAsnAsnSerAla 290
 1266 GGGTAAGCGGAATATGGCGGTGTATTCCTCAAAATCAGCAGCTA 1315
 290 ys.....IleGlyGlyAlaIleAlaIleGlnSerGly 300
 1316 AAGCGCGCAAGCTGATGATTACAGCGCATTAAGTCACATTAAACAGCT 1365
 301 GlySerLeuSerIleLeuAlaGlyGluGlySerValLeuPheGlnAsn 317
 1366 GCAGTTATCGACCTTCAGGTAAAGAGGAGGAACCTTACCTGCCGCG 1415
 317 nSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeu 334
 1416 TCACGAGCC.....GGCGAAGGTAAACGGCATTCATTAATG 1453
 334 LuLysAspAlaIleLeuSerSerLeuGluAlaArgAsn..... 346
 1454 CAAGCAAAACCTCTTTAGAAAAAGCTCACCATCATGTATCAGCAAA 1503
 347GlyAspIleLeuPhePheAspPr 354
 1504 GAAAAAGCGGAGCGCCATTTGTCGCGCGCATTTGCCCTTAATTCAC.. 1551
 354 oLleValGlnIleSerSerSerLysGlySerProLeuProSerSerLeuG 371
 1552GGCAATATTAA 1561
 371 InAlaSerValThrSerProThrProAlaThrAlaSerProLeuValIle 387
 1562 ACCCTCAAGGTAGTGTGATATCGCTAAACCGGTGT.....TTTGG 1605
 388 GlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerG 404
 1606 GAGACATCGCGCATTTATTCATGTACAGCAATGCCAATGTTAAAC 1655
 404 uGluGlu.....LysThrProAspAsnLeuThrSerGlnLeuGlnIle 419
 1656 AAAGAGTGTGTTGCTAGACCTGATGATGTAAACATTGAGCGGAGAC 1705
 419 rIleGlnLeuLysSerGlyArgLeuValLeuLysAspArg..... 432
 1706 CCTTCGCAATATACCGT.....ATTAATGATGAATTCACCAACA 1746
 433AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAla 447
 1747 GCGACCGGTGAAGCAAGCGACCTTAATAAAATAGCAACTCAAAACAC 1796
 447 uLeuIle..... 449
 1797 GCTAACCAATACAACTATTCTCAAAATATCTGAAAAACCGCTGGACAA 1846
 450 ..MetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAla 465
 1847 ATATTAACCGCATCAAGAAACTTACCTTAATACCTCAATCAACATCGCA 1896
 466 ThrLeuSer...IleProLeuHisSer..... 473
 1897 ACGAATCCCACTTAATTCATAGTAAGGTCAGCGTGGCGGAGCGCT 1946
 473 473
 1947 TCAGATTGATGAGATATTACTCTTAAGCGGAATTTAACCATTTATT 1996
 474LeuAspThrGluLysSerValThrIleHisAlaProAsn 486
 1997 CTGCGGATGGTGTGATCTCATAAATATTACCTTCAGTCAGCGGTTT 2046
 487 LeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspGluAsnPheTy 503

2047 TTAATATTACCGCGCTTCCTACCTTTTGAAGTGAATAATACAAAGC 2096
 503 rGlu.....AsnValGluLeuLeuSerLys..... 511
 2097 ACGCGAGCGGCAAAATGCTAAATATTGTCCGCGGAGCATGTACCATTA 2146
 512GluGlnAsnAsnIleProLeu..... 518
 2147 CAGGAGAGGAAACATTTACAGCGCTACACACATCTTTAAACGCAAGC 2196
 519LeuThrLeuSerLysGlnGlnSerHisLeuHisLeuProAs 532
 2197 GGTAAAGGTGTAATATCATTTTCATCAGTGAATATTAA.....AC 2237
 532 pGlyAsnLeuSerSerHisPheGlyTyrGlnGlnLysPhePheSer 548
 2238 CCACATCTTAGTGGCACAAATTACATATCTGGGAATATACAAATTAAC 2287
 549TrpLysAspSerAspGluGlnLys 556
 2288 AACCTACGAGAAAGAACACCTGCTATTGGCAACACGACATGATTCGCAC 2337
 557 SerLeuIle.....AlaAsnTrpThrProLys 565
 2338 TGGAGCTGAGTGTCTTAACTAGACAGCGCAAAATTTACCTTAT 2387
 565 sAsnTrpValProHisProGluArgGlnSerThr..... 576
 2388 TAAATACATTTCAAGCAATAGCAAAAGCTTAACACACAGATAGAGACT 2437
 576 576
 2438 CTGCAAGGGTGAATTTTAACGCGGTAATGGCAACATGTCATTCAATCTC 2487
 577LeuValAlaAsnThrLeuTrpAs 584
 2488 AAGAGAGGAGCGAAAGTAAATTTCAATTAACCAAGCAAGACATGAA 2537
 584 nThrTyrSerAspMetGlnAlaValGlnSerMetIleAsnThrIleAla 601
 2538 CACA...AGCAACCTTTACCAATTCGGTTTTACCAATATCAGAGCA 2584
 601 IsGlyAlaLysIleuPheGlyThrTrpGlySerAlaValSerAsnLeu 617
 2585 CTGCTGGC.....GCCTCGTCTTTTTCATATTA 2613
 618 PheTyrAlaHisAspSerSerGlyLysProIleAsp...AsnTrpHisHI 633
 2614 TATGCCAACCATTTCTGCGCAGAGGCGCTGAGTTAAATGATGAATTA 2663
 633 sArgSerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAsn 650
 2664 TATCTTAACGCGCTAATTTTACCTTAATTCATTCATGCTGGCGGATG 2713
 650 IsserPheCysLeuAlaIleGlnLeuLeuGlyLysSerSerAspSer 666
 2714 ACCGTTTAAATC.....ACAAACACTTA 2739
 667 PheIleThrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaG 683
 2740 ACCATTAATGCAACCATTA..... 2760
 683 nLeuAlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrAsnGln 700
 2761AATTTC 2767
 700 erIleHisGluLeuLysThrLys...TyrArgSerPheSerLysGlnGly 715
 2768 CCGTCAGACAGCAAGATGATGATTTTATGACGGGTACGACGCAATGCC 2817
 716 PheGlySerTrpHisSerValAlaValSerGly 726

```
117 SerProSerIleIleHisGlnLysAsnGlnLeuSerLeuArgAsnAs 133
    ::::: ||| ::::: ||| ::::: ||| ::::: |||
928 GCGCTCGCTGAATAT...GTGAATACAGGTTAATATAGTGGTAACA 974
133 nGlySerMetSerPheCysArgAsnHisAlaGluGlySerGlyAlaI 150
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
975 CGGCACTGTAATCTTATP...CGGGCAACA 1003
150 LeSerAlaSpAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
    :::::
1004 TGAA... 1008
167 GluGluAsnSerSerLysGlnGlyAlaIleGln...Al 180
    ::::: ||| ::::: ||| ::::: ||| ::::: |||
1009 AACGAGGCGTGTATAGCTTAATGCTGGCAGCATTTCTTACTCCAGG 1058
180 acGlnThrPheSerLeuSerArgAsnValSerPro...I 192
    ||||| ||||| ||||| ||||| ||||| |||||
1059 GCAAAAAACACCATCAGCGATATATAACCCACCATTTACTTACACCA 1108
192 LeSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCys 208
    ||||| ||||| ||||| ||||| ||||| |||||
1109 TTGCGCGCGCTTAATAATGAGCG...GTCAATCTGGCGATATTTTGGC 1155
209 SerAsnLeuIleCysSerGlyAsnValAsnProLeuPhePheThrGlyAs 225
    ::::: ||| ::::: ||| ::::: ||| ::::: |||
1156 AAA...GGCGTAACATTAATGTC...CG 1178
225 nSerAlaThrAsnGlyGlyAlaIleCysCysIleSerAspLeuAsnThrS 242
    ::::: |||||
1179 TCCTCCACT... 1188
242 erGluLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAla 258
    ||||| ||||| ||||| |||||
1189 ...ATTCAACACCAAGTAACCTTCTGCT 1215
259 SerAsnSerAlaLysGlnLysGlyAlaIleTyr...AlaLysHsI 273
    ::::: ||||| ||||| ||||| |||||
1216 GATCTGTACGCAAGATAAAGCGCAATATTGTTCTTCCGCCAAGA 1265
273 sMetValLeuArgTyrAsnGlyProValSerPheIleAsnAsnSerAlaL 290
1266 GCGTCAACCGGAATTTGGCGGTGTAATTTCCCTCAAAATCAGCAAGCA 1315
290 ys...IleGlyAlaIleAlaIleGlnSerGly 300
    ||| |||
1316 AAGCGGCAAGCTGATGATTACAGCGCATTAAGTACATTAACACAGCT 1365
301 GlySerLeuSerIleLeuAlaGlyGlnGlySerValLeuPheGlnAsnAs 317
    ::::: ||||| ||||| ||||| |||||
1366 GCAGTTATCGACCTTTCAGTAAAGAGGCGAGAACTTACCTTGGCGG 1415
317 nSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeuG 334
    ||||| ||||| ||||| ||||| ||||| |||||
1416 TGACAGCGC...GGCGAAGTAAAGAACGCAATTCATTAG 1453
334 LuLysAspAlaIleLeuSerSerLeuGlnAlaArgAsn... 346
    ||| |||
1454 CAAGAAACCTCTTTAGAAAAAGGCTCAACCATCAATGTATCAGGCAAA 1503
347 ...GlyAspIleLeuPhePheAspPr 354
    ||||| ||||| ||||| |||||
1504 GAAAGGCGAGCGGCTATTGTGGGGCATATTGCGTTAATGAC... 1551
354 oLleValGlnIleuSerSerLysGlnSerProLeuProSerSerLeuG 371
    ::::: ||||| ||||| ||||| |||||
1552 ...GGCAATATTA 1561
371 LuAlaSerValThrSerProThrProAlaThrAlaSerProLeuValIle 387
    ::||| ||||| ||||| |||||
1562 ACGCTCAAGTAGTGGTATCGTAAACCGGTGT...TTTGTG 1605
388 GlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerGI 404
```

```
1606 GAGACATCGGGCGATATTATTCATTCAGCAAGCAATTCATTGTTAAAC 1655
404 uGluGlu...LysThrProAspAsnLeuThrSerGlnLeuGlnP 419
    ::||| ||||| ||||| ||||| ||||| |||||
1656 AAAAGAGTGTTGCTAGACCCGTGATGATGATCAATATGAAGCGAAGAC 1705
419 rolleGluLeuLysSerGlyArgLeuValLeuLysAspArg... 432
    ||||| ||||| ||||| ||||| ||||| |||||
1706 CCCTTCGCAATATACCGGT...ATAATGATGAAATTCGCCACA 1746
433 ...AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaIle 447
    ::||| ||||| ||||| ||||| ||||| |||||
1747 GGCACCGGTCAGCAAGCACCCCTAATAAATAGCAACTCAAAACAC 1796
447 uLeuIle... 449
    |||
1797 GCTAACCAATACACTATTTCAAATATCTGAATAAGCGCTGACAAATGA 1846
450 ...MetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAla 465
    ::||| ||||| ||||| ||||| ||||| |||||
1847 ATATAAGGCAATCAAGAAACCTTACCGTTAATAGCTCAATCAACATCGGA 1896
466 ThrLeuSer...IleProLeuHisSer... 473
    ::||| ||||| ||||| ||||| ||||| |||||
1897 ACCAATCCCACTTAATCTCCATAGTAAGGTCAAGCTGCGGAGCGCT 1946
473 ... 473
1947 TCAGATTGATGAGATATTACTTCTAAACCCGCAATTTAACCATTTATT 1996
474 ...LeuAspThrGluLysSerValThrIleHisAlaProAsn 486
    ::||| ||||| ||||| ||||| ||||| |||||
1997 CTGGCGGATGGGTGATGATCTTAATAAATATTCCTGATCAGGGTATT 2046
487 LeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspGluAsnPheTy 503
    ||||| ||||| ||||| ||||| ||||| |||||
2047 TTAATATTACCGCGCGCTCCGTACCTTTTGAAAGTGCATAATACAAAGC 2096
503 rGlu...AsnValGluLeuLeuSerLys... 511
    ::||| ||||| ||||| ||||| ||||| |||||
2097 ACCGACGCGGCAAAATGCTAAATTTGCGCCAGCGCACTGTAAACATTAA 2146
512 ...GlnGlnAsnAsnIleProLeu... 518
    ||| |||
2147 CAGGAGAGGAAAGATTCAGGGCTTAACACGATCTTTAAACGGAAG 2196
519 ...LeuThrLeuSerLysGlnGlnSerHisLeuHisLeuProAs 532
    ||||| ||||| ||||| ||||| ||||| |||||
2197 GGTAAAGGCTGATATCATTTTCATCAGTGAATATTTA...AC 2237
532 pGlyAsnLeuSerSerHisPheGlyTyrGlnGlyAspThrPheSer... 548
    ||||| ||||| ||||| ||||| ||||| |||||
2238 CCACATCTTACTGTGACAAATTAACATATCTGCGAATATACAAATTAAC 2287
549 ...TrpLysAspSerAspGlnGlyHis 556
    ::||| ||||| ||||| ||||| ||||| |||||
2288 AAATCAGCAAGAAACACCTCGTATTGCAACCCAGCATGATTCGAC 2337
557 SerLeuIle...AlaAsnTrpThrProLys 565
    ::||| ||||| ||||| ||||| ||||| |||||
2338 TGGAACGTCACATGCTTATATCTAGACAGACAGCCCAATTTTACCTTAT 2387
565 sAsnTyrValProHisProGluArgGlnSerThr... 576
    ::||| ||||| ||||| ||||| ||||| |||||
2388 TAAATACATTTCAAGCAATACGAAAGCTTAACAACAGATAGAAAGCT 2437
576 ... 576
2438 CTGACAGGCTAATTTTACGCGCTAAATGCAACATGTCTTAATCTC 2487
577 ...LeuValAlaAsnThrLeuTyrAs 584
    ||| |||
```



```
784 ylllethphecJlulysSercIn...LysThrArgThrTyrTyrP 800
      ||||| |||:|||||:|||||: |||||
23567 GATCCGATTGTAGTAAGAAATCAAGACTGCCAAGATGCAAGTACATCTAA 23616
      ||||| |||:|||||:|||||: |||||
800 heLeuGIYAlaTYrIIeGIAspLeuLysArgAspValGIuSerGIyPro 816
      ||||| |||:|||||:|||||: |||||
23617 CTCCTTGGT...TATACGTGGAATCTTGTCTGCTAGTAACCCCGACTGTACG 23663
      ||||| |||:|||||:|||||: |||||
817 ValValLeuLeuLysAsnAlaValSerTRpAspAlaPrometaIAsnLe 833
      ||||| |||:|||||:|||||: |||||
23664 ACAACACTCGCAATTAACCGCTGATTTGCAAAACTTTCGCTACGAATTT 23713
      ||||| |||:|||||:|||||: |||||
833 uAspSerArgAlaTYrMetPheArgLeuThrAsnGlnArgAlaLeuHis. 849
      ||||| |||:|||||:|||||: |||||
23714 GGCAGACACAAGCTTTAGCTCCTCGTCAGGGAACCAATTTTGTCTTAAGT 23763
      ||||| |||:|||||:|||||: |||||
850 . ArgLeuGIInThrLeuAsnValSerCysValLeuArgGIyGlnSer 865
      ||||| |||:|||||:|||||: |||||
23764 CAATTTTGAAGCCTTAGCCCAATTTTCTTTTGAATTCGCTGGCTCATCT 23813
      ||||| |||:|||||:|||||: |||||
866 HisSerTYrSerLeuAspLeuGIYThrThrTYrArgPhe 878
      ||||| |||:|||||:|||||: |||||
23814 CGCAATTACATGTAGACTTAGAGCAAAATACCAATTC 23852
```

gene
CDS
4127..6895
/gene="pmp_1"
4127..6895
/gene="pmp_1"
/codon_start=1
/transl_table=1
/product="polymorphic outer membrane protein G family"
/protein_id="BA08215.1"
/db_xref="GI:8978378"
/transl_table="MRSFGPLVFSEFTLLSVEDTSLSATITSLPDEDSFHDSONA
ERSYVWQGDVSLTGIDVTSISVNDNSALNKAFVNVGSGYFAFNHGLVFNINSGDT
TKEGVALGQDPOAATFARPSGFTSLFIOSPDILKEOGCLYNALMLNNVYHEDON
OSKTKGASIGSANDYIVGNVDSFYFOAAATFGCAHSSGQVLAVNQAARFRAONRA
KNSGALSGSGDIDIDONAVLFRENDALTATAGKGNCCDPTSSGSRPAPVATES
DNKOLVFRNNISIMGGCAITARKLSISGGFTLTINNI STANSONLGGALAIIDTGET
SLSAKGTITTOGNRTSLPFLNGIHLLQNAFETLQNRNGSIEFYDPTISEADGSGTO
LNINDPKNKITYCTTILPFGSKSLANDPREFKSTPONVINLSAGYLIKKCAEYVATSK
FTOSGSLVIDLDCIKLASKEDIAITGIALIDLSLSSSTAAVIRKANTKANKOISVTD
SILISLPGNVAEFDLRMKNISOTFPLLSLEPGACGSGYVATGDAPLPSAPHYGPGNNKL
AMTGKAKVGEFENDKIIYKRPKECHLVPNIIKGNADVRSIMOVETHASSLOTD
RGLVTDIGNFTFIVSASEDNIYRHNSSGVLVYNNNETTPKHYTSMAFSOLFSDKQY
AVSNNEYMYGSLYLOYTTSLGNI FRVANSNPVNGIISLRFLOMPLEHIFELCY
CHATDNKMTDPAFNPVNNSWRNKNCALIECGSSPLLVFENGRLFOCALIPMKLOLVY
AYOGGFKETPDADKRFNSGLSTISVPLGIFEKALISODVLYGFTFSYPIIDFRKQY
SCEALVIVSGDSMLVPRAHVSRAHAFVGGTGTRKHFNDYTLLCGSIECHRHANNYNI
NCGSFR"

gene
CDS
7515..10169
/gene="CPJ0007"
7515..10169
/gene="CPJ0007"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BA098216.1"
/db_xref="GI:8978379"
/translation="MQEPLRSALLERLSEWLVLLOVSPETTRSTPEKDANQLPKDSK
NPVLESL"

gene
CDS
7515..10169
/gene="CPJ0007"
7515..10169
/gene="CPJ0007"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BA098217.2"
/db_xref="GI:9956083"
/translation="MFSFLVIVPLTDSTTSLSLMLDEGNQPMRKRLRIYALIAL
SILIAAGVILLTVAIPGLSSVISPAGMGALCGVLAAGTIOVLKKEPEVIVLAS
VTTTPGIVSPSGISISGADSTISLPYLLDECHPOSMKRLIAVILVFSIIILIA
SGVLLTVAIPGLSSVISPAGMGALCGVLAAGTIOVLKKEPEVIVPAPPEEV
VIDIDELTIRLOOEAENALARPPEEMSAFEGYIKGIVLHNNKSLPYDPIHGLEET
KHQIVVSSLSKAMPPEFLDIRIIFEEFEFFLSARRKLIIDATLVKRLITFOLUR
NNLRKAFSVLODSIFPKIINDNPELAKPMILSKSICRPTIIFENNEHGVASLHKH
NAVILEKIVRSIKVSDYDIOGSSAKKMLIHLGNPFSLDENKKITMKHEHMLIESLSS
YRKVLAIDSENVDTSPDPRKKWDSGIPCDALSELISRDQWKKAKLKHQESLYTIO
ARADLTDOSSKENOELEKAEQYIISWEKAKKEIERVORIDAIQKLVPMILEREKE
ETTGGETVTPYVQGTASSDLIDLIGRLEVSRBDNONOESCVALVRSHEVMSWEEK
OEYGGKKEFDDOMKSLERFTEHIELELVODKDYKHLVYFKVNNKKEYOYAKFEL
KVLSEDLGLIAOTESASLITJOELPLIATRGALKEAIVGSGISALASAKRYFEE
DPRDSDSTPRLATLRILOEAKASLEIEIKFNSLNDIAFERRLKESKOTPRALDI
GVLRINVESTYDILRSLNTWEGTPESKRYVFSMYLNTYNEEKRAKTRLVEMTORIR
DFKMLEMQRNEELLOEELSLQAPSE"

gene
CDS
10645..11358

[illegible]

Align seg 1/1 to reverse of: AE002235 from: 1 to: 12127

```

126 GlyClnLeuSerLeuArgAsnAsnGlySerMetSerPheCysArgAsnI 142
1673 GGGTACCTGGAATTTGACAAAAATGCGATTGGCTTCCACCAAAACTT 8624
142 sAlaGluGlySerGlyGlyAlaIleSerAlaAspIlePheSerLeuGlnI 159
8623 TTCACAGGATATATGGCGCGTCATATCCGCCCAAAACCTCTTTCATTAC 8574
159 IAsnGlyLeuPheThrAlaPheGluAsnSerSerLeuGlyAsnGly 175
8573 GGACTCAATGTCACCTCTGTTTCTGAAAAATACCTCCTCAAGAAAGCG 8524
176 GlyAlaIleGln...AlaGlnThrPheSerLeuSerArgAsnValSerP 191
8523 GGGAGCTTACACACTTCGCATGCCCTTCCATCTACATTCGAAACCAAGG 8474
191 oIleSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCysC 208
8473 AGCTGCTTTTTCGACAAATACCTCTTCGGATTCTGGACCTGCAATTTTGA 8424
208 ySerAsnLeuIleCysSerGlyAsnValAsnProLeuPhePheThrGly 222
8423 CAGAACCTCGGTGACTATTCTATATATGCTAAAGTTCTTATTATGAC 8374
225 AsnSerAlaThr.....AsnGlyC 231
8373 AATTAAGTTCACGAGCGAGCTCTCAACACGGGGATATGTCAAGAGG 8324
231 yAlaIleCysCysIleSerAspLeuAsnThrSerGluLeuGlySerLeuS 248
8323 TCGTATCTGTCT.....TTATAAACCTAGTACACATACATAAGCTCA 8283
248 eLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaGlyGlu 264
8282 CCTCTACTGGAATCGATGTACTACTTCACACACAAATACATCAACAACA 8233
265 LysGlyGlyAlaIleThrValAlaIshHisMetValLeuArgTyrAsnGlyP 281
8232 GCGGGAGGAGCTATCTATGTGAAAAAGCTCGACCTGGCTCCGAGGACT 8183
281 oValSerPheIleAsnAsnSer.....AlaLysIleGlyC 293
8182 TTACCATTAATCAGTAACATAGTCATCAAGCAACATCCATCTCAAGGTC 8133
293 LysAlaIleAlaIleGlnSerGlyGlySerLeuSerIleLeuAlaGlyGlu 309
8132 GAGCCATACCTATCGAAGATATAGTGGGATGAGTTTATCCGCCGATAGT 8083
310 GlySerValLeuPheGlnAsnAsnSerGlnArgThrSerAspGlnGlyLe 326
8082 GGTGACATGTCTTTTATGGGATACATCACTCTTACTACTCTCTGGGC 8033
326 uValArgAsnAlaIleTyrLeuGlnLysAspAlaIleLeuSerSerLeuG 343
8032 GAATAGAACTACTATCGACTTGAAGAACAGATGCCAAAGATGACACTTTGC 7993
343 LysAlaArgAsnGly...AspIleLeuPhePheAspProIleValGlnGlu 358
7993 GTTCTGCTGCTGCTGAGAGCCATCTCTATGTATCCCATACATCAAGGA 7933
358 SerSerSer.....LysGlnSerProLeuP 367
7932 TCATTCACACAGCTTACAGATGTCTTAAAGCTTATATGAGACTCCGGCAGA 7883
367 oSerSerLeuGlnAlaSerValThrSerProThrProIleThrAlaSerP 384
7882 TTCTGCACATCAAAATATACA.....7864
384 roleuValIleGlnThrSerAlaAsnArgSerValIlePheSerSerGlu 400

```

7863GGGACATCATCTTCCAGAGGACA	7840
401	ArgLeuSerGluGluGlyuSerThrProAsp...AsnLeuThrSerGlnLe	416
::: 	
7839	AASTPTATACAGACACAGAGCCGACAGATTCTAAAAATCTTACTGGAACT	7790
416	uGlnGlnProIleGluLeuLysSerGlyArgLeuValLeuLysAspArg	433
	
7789	ACTACAGCCCTGAACCTCTTTCAGAGAGTACCTTACTTTAAACATGAGC	7740
433	IaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeuLeuLe	449
	
7739	TGACTGTGCAGACTCAGGCATTCTACCAACAGCAGATTCCTGCTCGCA	7690
450	MetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAlaTh	466
	
7689	ATGAGCTAGGAACCTAGCTTACAGACCTCTCTACTACACCATTAACA	7640
466	rLeuSerLleProLeuLHisSerLeuAspThrGluLysSerValThrIleH	483
	
7639	TTTGGCTATTAACATCCTACTCTATGACCGCTGCAAGACGAAATAAG	7590
483	Is.....AlaProAsnLeuSerLleGlnLysLlePheLeuSer	495
	
7589	AAACCAACAGCTACGTGTAATAAAATCTGACTTTATCTCGAACCATCTT	7540
496	AsnSerGlyAspGluAsnPheArgLysAsnValGluLeuLeuSerLysG	512
	
7539	TTCGACCCGACGCGGCGCGCTTTATGTAATAATCTATTAAGAATCTTCA	7490
512	uGlnAsnAsnLleProLeuLeuThrLeuSer...LysGluGlnSerHisL	528
	
7489	GTTCCTACGACATCTTGAAGCTCAAACTCTTCGAACTGTAAACAAGACCG	7440
528	eunHisLeuProAspGluLysnLeuSerSer...HisPheGlyTyrGln	542
	
7439	CAGTACTCCCAATCTCTAATAAGGCTGAAGAAATCTCATACGCTATCAG	7390
543	GlyAspTrp...ThrPheSerTrpLysAspSerAspGluGlnHisSerLe	558
	
7389	GGAACTGGGGCCCAATGTCTTGG.....GGCAGAGGGCGTCTTAC	7349
558	uIleAla.....AsnTrpThrProLysAsnTyrValProHisProGluA	573
	
7348	GAGTGCACACCTTCAACTGCAGCTAAACACCTGCATATCTCAATCCGACG	7299
573	rgGlnSerThrIleuValAlaAsnThrLeuThrProAsnThrTyrSerAspMet	589
	
7298	GTAATCGCTCTTATGCTCCCTAATACCTTATGGAAAGCATTTATAGAAATT	7249
590	GlnAlaValaGlnSerMetLleAsnThrIleAlaHisGly.....GlyAl	604
	
7248	AGCTCTCTCCATTTCTTATGAGACGTCAAAACGAGGGTTCCAGCGCAGA	7199
604	ArgLeuPheGlyThrTrpGlySerAlaValSerAsnLeuPheTyrLanH	621
	
7198	CCGTGCTTT.....TGCTGCTGGATTATCTACTTCTTC.....C	7161
621	IAspSerSerGlyLysProIleAspAsnTrpHisHisArgSerLeuGly	637
	
7160	ATAAGAGATGTGCAAAACACAGCAGCGGGTTCTCGCATTTAGAGCGCGT	7111
638	TyrLeuPheGlyIleSerThrIleSerLeuAspAsnHisSerPheCysLe	654
	
7110	TATGTCTATAGGAGAAACCTACATCTTGTTCAGTAAAGATTCTTACTGC	7061
654	uAlaAlaGlyGlnLeuLeuGlyLysSerSerAspSerPheIleThrSer	671
	
7060	TGCATTGTTCTAGCTCTTTGGAAAGAGATAGAGAGCTACTTGTACTAGAA	7011
671	hGluThrThrSerTyrIleAlaThrValGlnAlaGlnLeuAlaThrSer	687
	
7010	ATCAAGGTACAGCTACAGAGAGAGATCTCTATTACACGACACAAAGCAAC	6961

```
3524 CAGCTTATGTTAGATTCCTATCTAACATAGCGTCTACGTCAGAGCGCCT 3573
206 IleCysSerSerLeuIleCysSerGlyValAsnProLeuPhePhe 222
3574 ATCATGTGTGAAGCAGCTCGATCTACTATCGAACACAAATTTCTATATT 3623
222 ethGlyAsnSerAla...ThrasnglyAlaIleCysCysIleSera 238
3624 TGAAGGGAAATCGACGAAACTACTGGCGGTGCATCTGC..... 3663
238 sPLeuAsnThrSerGluSerGlySer.....LeuSerLeuAlaCysAsn 252
3664 ....AACCCAGCGAGCTGGATTCCTGTAACGATATCTCTCAACAAAT 3708
253 GlnGluThrLeuPheAlaSerAsnSerAlaLeuSerGlyValAlaIle 269
3709 AAGACTCGATCTTGGCTTCAACGTAGCAGAACAAAGCGGTGGCGCAT 3758
269 eTyAlaLysHisMetValLeuArgTyrAsnGlyProValSerPheIle 286
3759 CCATGCTTAAAGCTAGCCCTTCTCTGAGGCTTACAGAGTTTCTAC 3808
286 snAsnSerAlaLysIle.....GlyGlyAlaIleAlaIleGln 298
3809 GAAATTAATGCTCATCAGCAACTCCTTAAGCGGCTGCTATCAGCATCAT 3858
299 SerGlyGlySerLeuSerIleLeuAlaGlyGluGlySerValLeuPheG 315
3859 GCCTCAGAGAGCTCAGTCTTCTTGCAGACAGACAGAAACATTACTTTGT 3908
315 snAsnSerGlnArgThr.....SerAspGlnGlyLeuValArgAsn 330
3909 AAGAAATACCTTACAAACACCGGAAATGACCTCTCAAGCATATG 3958
330 laIleTyrLeuGluLysAspAlaIleLeuSerSerLeuGlu...AlaArg 345
3959 CGATCAACATAGCAAGTAAACGGAAATTCACCGAAATTCACCGCTCCTAA 4008
346 AsnGlyAspIleLeuPhePheAspProIleValGlnIleSerSerSer 362
4009 AATCATCAATTTTCTTCTATGATCCCTACCTTCAGAGAACCTCATC 4038
362 sGluSerProLeuProSerSerIleuGlnAlaSerValThrSerProThr 379
4059 AGAGTATTGAAGATAAATACGGCTCGCGAGCTCTCAATCATCATTC 4108
379 roAlaThrAlaSerProLeuValIleGlnThrSerAlaAsnArgSerVal 395
4109 AAGGACG.....ATT 4119
396 IlePheSerSerGluArgLeuSerGluGlu...LysThrProAsp 411
4120 CTATTTTGGAGAAACCTTAACAGACATGATCTAAAGTTGCGACAA 4169
411 nLeuThrSerGlnLeuGlnProIleGluLeuLysSerGlyArgLeu 428
4170 TTTAAATCTTCATTCACGACGAGTCTCTCCATCCGAGAAAGTTAT 4219
428 alLeuLysAspArgAlaValLeuSerAlaProSerLeuSerGlnAsp 444
4220 TCCTACAAAGAGAGTCACTTACAGACACACGCTTCTCAAGGCC 4269
445 GlnAlaLeuLeuIleMetGluAlaGlyThrSerLeuLysThrSerSer 460
4270 GGTTCCTCCCTCGCATGATTCAGAGACGACATTATCACTACAGCTGG 4319
461 AspLeuLysLeuAlaThrLeuSerIleProLeuHisSerLeuAspThr 477
4320 GAGTTTACATCAACGACCTAGCAATCAATGATCTTACCTTAGCTTAA 4369
477 LuLysSerValThrIleHisAlaProAsnLeuSerIleGlnLysIlePhe 493
4370 AGCAGCCCTCAGCCTAACAGCAAAAGCTCTTCA...AATAACATCATC 4416
494 LeuSerAsn.....SerGlyAspGlnAsnPheArgLys 505
4417 GTATCTGGGAAGCTCAACCTGATGATATTCAGAGAACATTATGAAAG 4466
505 nValGluLeuSerLysGluGln..... 513
4467 T...CATATGCTTACAGCCATGACACAGCTCTCTCTATTTAAATTCACGG 4513
514 .....AsnAsnIleProLeuLeuThrLeuSerLysGlu 524
4514 TTGATGCTGATGTTGATATCAATCACTGACATCAACACCTT..... 4554
525 GlnSerHisLeuHisLeuProAspGlyAsnLeuSerSerHisPheGly 541
4555 .....ATCCCTGTTCTCGCTGAGAGATCTTAATCAGATACGAT 4595
541 rGlnGlyAspThrPheSerThrPheSerLysAspSerAspGluLysSer 557
4596 CCAAGACAAATGCAATGTAAATGAGTACAGGAAACAGTACAAATACAA 4645
558 .....LeuIleAlaAsnThrPheProLysAsnTyrValProHisProGlu 572
4646 AAGAGCCACCGCAACTTGGACCAAAACAGAGATTTGTCACGCCCGAA 4695
573 ArgGlnSerThrLeuValAlaAsnThrLeuThrPheAsnThrTyrSerAsp 589
4696 AAGAAATCTGCGTATGATGCAATACCTATGGGAGTCTTACTGACAT 4745
589 rGlnAlaValGlnSerMetIleAsnThrIleAlaHisGlyAlaTyrL 606
4746 TCCTCTCTGCAACAGCTTGTAGAGATCGCCCAACTGTTCGACACACA 4795
606 ePheGlyThrTyrGlySerAlaValSerAsnLeuPheTyrAlaHisAsp 622
4796 AACAGGTTCTTGGTTCCTCATGACGAAC.....TTCCTCAACAAAG 4839
623 SerSerGlyLysProIleAspAsnThrHisArgSerLeuGlyTyrLe 639
4840 ACTGAGATGAAATCCGAAGGCTTCGCTATACCTTGGAGGCTACGT 4889
639 uPheGlyIleSerThrHisSerLeuAspAspHisSerPheCysLeuAla 656
4890 CATCGTGGAAGTCTGACACTCTAAAGACGACCTATTACCTTTCGCT 4939
656 laGlyGlnLeuLeuGlyLysSerSerAspPheIleThrSerThrGlu 672
4940 TCTGCCATCTTGTGCTAGACACAAGATGTTTATCGCTCACACACAC 4989
673 ThrThrSerTyrIleAlaThrVal..... 680
4990 TCTAGAACCTACGGTGGACATTATTTCTCAAGCAGCTCTCATACCCTACA 5039
681 .....GlnAlaGlnLeuAlaThrSerLeu 689
5040 ACCCCAAACTATTGAGATTAGAGAGCAAGTTTCTGAATCAGCTA 5089
689 eLys.....IleSerAlaGlnAlaCys 696
5090 TAGAAAAATTCCTAGGAAATTCCTCAGCTTGGAGTGCAGATTCG 5139
697 TyrAsnGlnSerIleHisGluLeuLysThrLysTyrArgSerPheSer 713
5140 TTCAGCCATTAGAACACCGATAGGAAGCAGCATATACCTGATTCG...CC 5186
713 sGluGlyPheGlySerThrPheHisSerValAlaValSerGlyValLys 730
5187 AGAATCCCAAGCTTCTGACACACAGAGTGTATAGCTGTGCTATCCGCC 5236
730 laSerIlePro...IleValSerAsnGlySerGlyLeuPheSerSerPhe 745
5237 TAGACCTTCTTCTTGTCTTTCACCAACCCACATCTCTTTCAAGACCTTC 5286
```

```

746 SerL1epheserLyLeuGlnGlyIhpheserGlyThrGlnAspGlyPhagI 762
2587 ATTCACAGATGAAAGTCCAAATGGTTATATGATATACCAAAATAGCTTCT 2636
762 uGluSerGlyGluIleArgSerPheSerAlaSerPheArgAsnI 779
2637 CGAAAGCTCTACTGATGGCCGTGGTTTATGATATGGAAGCGTGTAAAC 2686
779 LeSerLeuProIleGlyIleThrPheGlnLyLySerGlnLyThrArg 795
2687 TCTCGATTCCTCTGGCGGCAAAATTCGTCACAGGGGCAATATCGAGATTTC 2736
796 ThrTyrTyrTyrPheLeuGlyValATyrIleGlnAspLeuLyArgAspA 812
2737 TACACCTATGATCTCTCAGAGATCTTTGTTCCGATGCTATACGTAAACA 2786
812 IGluSerGlyProValIValLeuLeuLyAsnAlaValSerThrAspAlaP 829
2787 TCCCAATCTACACGACGACTCTGTGATGATGCCACAGACTCTTGAAAAATTC 2836
829 rOmEAlaAsnLeuAspSerATGAlATyMePheArgGlyThrAsnGln 845
2837 CGCGTCCGCAATCTTTCACAGACGCACTTTTACTGACGGCGGTAGAAC... 2883
846 ArgAlaLeuAlaArgLeuGlnThrLeuLeuAsnValSerCys..... 859
2884 .....AAGTACGCTACACACTCCAAATGATGGAGCTCT 2915
860 .....ValIleuArgGlyGlnSerHisSerTyrSerLeuA 871
2916 CGGACATTCAGCTATGCAAGACTCCGCGATCTTCAGAGCACTCAAAATGAG 2965
871 sPLeuGlyThrThrTyrArgPhe 878
2966 ATGTGTGGTACCAACTCCGATTTC 2988

seq_name: gb_ba2:CPOMP54

seq_documentation_block:
LOCUS CPOMP54 6030 bp DNA BCT 22-JAN-1999
DEFINITION Chlamydia pneumoniae omp5 and omp4 genes.
ACCESSION AJ001311
VERSION AJ001311.1 GI:3255934
KEYWORDS omp4 gene; omp5 gene; outer membrane protein 4; outer membrane protein 5.
SOURCE
ORGANISM Chlamydia pneumoniae.
REFERENCE
AUTHORS Chlamydia pneumoniae
TITLE Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
JOURNAL Knudsen,K.
PUBLISHED Direct Submission
SUBMITTER Submitted (28-AUG-1997) Knudsen K., Department of Medical Microbiology and Immunology, University of Aarhus, The Bartholin building, University of Aarhus, DK-8000 Aarhus C, DENMARK
AUTHORS 2 (bases 1 to 6030)
Knudsen,K., Madsen,A.S., Mygind,P., Christiansen,G. and Birkeland,S.
TITLE Identification of two novel genes encoding 97- to 99-kilodalton outer membrane proteins of Chlamydia pneumoniae
JOURNAL Infect. Immun. 67 (1), 375-383 (1999)
FEATURES
source
1. .6030
location/Qualifiers
/organism="Chlamydia pneumoniae"
/isolate="CDC/CWL-029/VR-1310"
/db_xref="taxon:83558"
/clone="pEX2-18"
/clone="B8F3"
/clone_1lb="pbluescript"
/clone_1lb="pEX2"
/dev_stage="elementary body"
complement(1..2804)

```

22128 AAGATGAGTACTTACTTCCAAATACTATACCGAGTCGAGGATCG 22177
 447 LeuLeuIleMetGluAlaGlyThrSerLeuLysThrSerSerAsp...Le 462
 22178 AAAGCTGTTATGATGAGGAGGACTCTTTGAGCGACGCGTGGAGGGGT 22227
 462 uLysLeuAlaThrLeuSerIleProLeuHisSerLeuAsp...ThrGlu 478
 22228 CACTCTCAATGGCTTACGCTTAATATGATCTTATAGTGGGAAATA 22277
 478 ySerVal.....ThrIleHisAlaProAsnLeuSerIleGlnLys 491
 22278 AAGCTATCATTAAGCGCGACGACCACTAAGCATCTTCCCTTATACGAGG 22327
 492 IlePheLeuSerAsnSerGlyAspGluAsnPheTyrgluAsnValGluLe 508
 22328 CCTATCATGCTTGTAGATGCTCAGGGGAACTATATAGCATCATAACT 22377
 508 uLeuSerLysGluGlnAsnAsnIleProLeuLeuThrLeuSerLysGlu 524
 22378 C.....AGTCACAGCAGGCTTCTCTTAAATAGAGCTTCTGACACAG 22421
 525 GluSerHisLeuHisLeuProAspGly.....AsnLeuSer 536
 22422 GAAGGATGACTACTACATATATCCCGCATACCCCAATTTCTAAATACTACG 22471
 537 SerHisPheGlyTyrglyGlyAspTrpThr.PheSerTrpLysSerAsp 553
 22472 AATCACTATGGGTATCAAGGAGACTGMAATATGTTGGTGGCAGCATG 22521
 553 spGluGlyHisSerLeuIleAla.....AsnTrpThrProLysAsnTrp 567
 22522 CAAGTCGAAAAAATAAATGTCTTAACTTGAAGCTTGAAGAAAGAGATC 22571
 568 ValProHisProGluArgGlnSerThrLeuValAlaAsnThrLeuTrpAs 584
 22572 AAGCGAATCCAGAACGTCAGCAGCCTTTGCTTCTTAATACCTGTGGGG 22621
 584 nThrTySerAspMetGlnAlaValGlnSerMetIleAsnThrIleAlaH 601
 22622 TTTCTTTGTCATGTCCGCTCATTCAGAGCCTCATGACGAGGACACAA 22671
 601 IsGlyGlyAlaTyrlLeuPheGlyThrTrpGlySerAlaValSerAsnLeu 617
 22672 GTTGTGATATCTTCCTCAACAAATTTGGGTATCGAATGCGGAC... 22718
 618 PheTyAlaHisAspSerSerGlyLysProIleAsnTrpHisHisAsr 634
 22719 TTTTTCATGTAAGATCAGAAAGCAACCA...CGTAACTTATCGTCATTC 22765
 634 gSerLeuGlyTyrlLeuPheGlyIleSerThrHisSerLeuAspAspLys 651
 22766 TAGCGGCGTTATGATTAGAGAGAGATTTCTGACGGCTCTGAAAT 22815
 651 erPheCysLeuAlaAlaGlyGlnLeuLeuGlyLysSerSerAspSerPhe 667
 22816 TCTTAAATTTGCTTTTGTGCTTTTGGCTTACGACAAAGACCACTT 22865
 668 IleThrSerThrGluThrThrSerTyrlIleAlaThrValGlnAlaGln 683
 22866 CTGGCTAAGAACCTACCATGTTATTCAGAGGCGCAATGACTTACCCACA 22915
 684LeuAlaThrSerLeuMetLysIle..... 691
 22916 CTTGAGAGAGCTTAAGACCCCTGCTAGATTGTCAGAAATTCGACT 22965
 692SerAlaGlnAlaCysTyrglnGluSerIleHis 702
 22966 CCTACCTTTTGTGTTCAATGCTGTGGTTTGTATTTGGCCATACGAGACAT 23015
 703 GluLeuLysThrLysTyrlArgSerPheSerLysGluGlyPheGlySerTr 719
 23016 AACATGACCAACAAAGTACACGCTATTTCTTCCTGTTAG...GGAACCTG 23062

719 phISerValAlaValSerGlyGluValCysAlaSerIleProIleValS 736
 23063 GGAAGATGATGCTTCGGTATAGAAATGTGGAGGACTTCCCGGAGTGG 23112
 736 erAsnGly...SerGlyLeuPheSerSerPheSerIlePheSerLysLeu 751
 23113 CTTCAGCAGCTGCGCTCTTGGTGCATACCCACACGCCATTTCTTAACCTA 23162
 752 GlnGlyPheSerGlyTyrlGlnAspGlyPheGluGlnSerSerGlyGluI 768
 23163 GAGATGATCTATGACATGAGAAATGACTTTAAGGAAACGCGACAGAGG 23212
 768 eArgSerPheSerAlaSerSerPheArgAsnIleSerLeuProIleGlyI 785
 23213 CGGTCTTCCAAAGTGAAGACCTTCTCAATCTACCGGTTCTCTGAGGA 23262
 785 leThrPheGluLysLysSerGlnLysThrArgThrTyrlTyrlPheLeu 801
 23263 TAAATTTGAGAAATTCCTCCATTAAGTCT...AGTATGATCTCTCCATA 23309
 802 GlyAlaTyrlIleGlnAspLeuLysArgAspValGlnSerGlyProValA 818
 23310 ...GCTTACGTTCCCGATGATGATTCCTAATGATCCAGGCTCCACGACAC 23356
 818 IleuLeuLysAsnAlaValSerTrpAspAlaProMetAlaAsnLeuAspS 835
 23357 TCTATGTTCTGCGGATTTCTGGTGCACATGAGTGCACAGCTTGCTTA 23406
 835 erArgAlaTyrlMetPheArgLeuThrAsnGlnArgAlaLeu...HisArg 850
 23407 GACAAAGCTCTTCTGTACGTCCTGGAATATCATATGCTTTCCTTCAAC 23456
 851 LeuGlnThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisS 867
 23457 TTTGACGTTTTCAGTCACTGTTGAAGTGCAGTTGCGAGGTTCTTCTGTA 23506
 867 rTyrlSerLeuAspLeuGlyTyrlThrTyrlArgPhe 878
 23507 CTATGCTATGATCTCTGGAGGAAGATTCGATTT 23540

seq.name: gb.pat1:A81827

seq.documentat_block:

LOCUS	A81827	DNA	PAT	21-JAN-2000
DEFINITION	Sequence 1 from Patent WO9858953.			
ACCESSION	A81827			
VERSION	A81827.1 GI:6731863			
KEYWORDS	unidentified.			
SOURCE	unidentified.			
ORGANISM	unclassified.			
REFERENCE	1 (bases 1 to 3200)			
AUTHORS	Madsen A. and BirkeLund S.			
TITLE	NOVEL SURFACE EXPRESSED PROTEINS FROM CHLAMYDIA PNEUMONIAE			
JOURNAL	PATENT: WO 9858953-A 1 30-DEC-1998;			
	MADSEN ANNA SOFIE (DK); BIRKElund SVEND (DK)			
FEATURES	Location/Qualifiers			
source	1..3200			
	/organism="unidentified"			
	/db_xref="taxon:32644"			
CDS	205..>2987			
	/note="unamed protein product"			
	/codon_start=1			
	/protein_id="CAB69199.1"			
	/db_xref="GI:6731864"			
	/translation="MKTSLIPWLVSSVLAFFSCHIOSLANELLSPDSEFNGLIDSGTF			
	TPKTSATVYSLNGDVEFPEPKGTPLSQCFKQTDNLFLGNGHSLTFGPIDAGCTHA			
	GAAASTYANKLTFSGFSLSDSPSTVTVTGGTGLSSAGVLENIKRLVYAGNFS			
	TADGGAIKASFLTGTSGLDALEFNNSSSTKGAIAATAGARIANNIGVRFSLNIA			
	TSGGAIDDEGTSLLSNKFLYPEGNAATKGTGALINTASGSPETLISNNKTLIFAS			
	VAEFGCAIHAKKKLALSSGTFPELRNNVSSATPFGAISIASELSLSAETGNITF			
	VAVPLTTGSTDTPKRNATINISNGKFTELRAKHHITTFYDPIITSEGTSSDVLKINN			

DEFINITION Chlamydia pneumoniae GGA1-protein family cluster B, genes omp5, hbl1, hb2, hb3, hb4, hb5, hb6, hb7, omp7, pseudogenes omp8, omp9, omp15, strain VR1310.

ACCESSION AJ133035

VERSION AJ133035.1 GI:4455890

KEYWORDS hbl1 gene; HB2 gene; hb3 gene; hb4 gene; hb5 gene; hb6 gene; hb7 gene; omp15 gene; omp6 gene; omp7 gene; omp8 gene; omp9 gene; outer membrane protein 6; outer membrane protein 7; pseudogene.

SOURCE Chlamydia pneumoniae.

ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

REFERENCE 1 (bases 1 to 26920)

AUTHORS Boesen, T.

JOURNAL Direct Submission

TITLE Submitted (21-JAN-1999) Boesen T., Department of Medical Microbiology and Immunology, University of Aarhus, The Bartholin Building, DK-8000 Aarhus, DENMARK

REFERENCE 2 (bases 1 to 26920)

AUTHORS Hjerno, K., Boesen, T., Dagaard, L., Knudsen, K., Madsen, A., Christiansen, G. and Birkelund, S.

TITLE Chlamydia proteins containing the GGA1-repeat belong to a subfamily of autotransporting pathogenicity factors

JOURNAL Unpublished

FEATURES

source Location/Qualifiers

1. 26920

/organism="Chlamydia pneumoniae"

/strain="VR1310"

/db_xref="taxon:83558"

498..575

/gene="omp6"

498..3266

/gene="omp6"

498..3266

/note="putative"

/codon_start=1

/transl_table=11

/product="outer membrane protein 6"

/protein_id="CAB37075.1"

/db_xref="GI:4455891"

/translation="MRFSLGCPFLVSLFLLSVFOTSLATTSITLPEDSPHGDSONA
FRSYVQADGVVSLNGDVLSISVNDNSALNKACFNVTSGSVTFAGNHGGLYFNNISGCT
TKRGAVLCODQDQATARFSGFSTLSYIQSGDIKROGCLYSKALMLNNYVRFQON
QSTKRGALISGANVTIVGNTDSYFOTNATFGALHSSGPLOIAVNOEIRFQNTA
KNQSGGALSDSDIDIDONAVLYFRENELTTAIGKGAVCCPTSGSSIPVPIVES
DNKQLEFERNHSMGGAIYARKLSISGQPTLFINNISYNSQNLGGAIADTGEI
SLSAQETITFOGNTSLPFLNGLHLLONAKFLKLDARNGCSIEFYDPIITSEADGSTO
LNINGDPKNEYGTITLFSGKSLANDPRDFTSTIONNLSAGYLVIKGAQVIVSK
FPOGSHLVLDCTKTLASKEDLATITGLAIDDISLSSSTAIVIKANTANNOISITD
STELISPTGNAYEDLIRMNSQTFPLLSLEGAGSVTVAGDFLPVSPHFGQGNKL
AMTGNKNKVEFEFMDKINYPREKGNLVPNLMGNANVRSLMQOETHASLQTD
RGLMIDGIGNFHVASEDNIRYRHNSGGVLSVNNETPKHTSMASFQLEFRDQY
AVSNNEYRMVLAGSYLYQYTTSLGNIYVARSNNVAVGLSRFLQNPIMIEFLCAY
GATNDMKITDYANFPMVKNSMRNCMAIECGSGMPLLVFENGRLPGCAIPFKLQLY
AYOGDFKETTADGRRFSGSLTISVPLGIRFEKLAISODVLYDFSTYIPDIFKDP
SCEAALVISGDSVLVPAHVSRAHFVSGTGRHFDYELLCRGSIIECRPIARANYNI
NCQSKFRF"

498..3266

/gene="omp6"

3976..6870

/gene="hb1"

3976..6870

/gene="hb1"

/note="hypothetical"

/codon_start=1

/transl_table=11

/product="Hbl protein"

/protein_id="CAB37076.1"

/db_xref="GI:4455892"

/translation="MRRRLIIVLIALSIILAGVLLTVAIPGLSSVISPACMG
ACALGCVMLALGIDVLKRRPILVASTTTPGSGPSGCSISGADSTIRSLPYL
LDGCHPSMRKLIIAIVLIVPSIILAGVLLTVAIPGLSSVISPACMGALCIC
VMALCIDVLKRRPILVASTTTPGSGPSGCSISGADSTIRSLPYLPDDEHP
OSMRKLIIAIVLIVPSIILAGVLLTVAIPGLSSVISPACMGALGCVMLALG
IDVLKRRPILVASTTTPGSGPSGCSISGADSTIRSLPYLPDDEHP

gene

498..3266

/gene="omp6"

3976..6870

/gene="hb1"

3976..6870

/gene="hb1"

/note="hypothetical"

/codon_start=1

/transl_table=11

/product="Hbl protein"

/protein_id="CAB37076.1"

/db_xref="GI:4455892"

/translation="MRRRLIIVLIALSIILAGVLLTVAIPGLSSVISPACMG
ACALGCVMLALGIDVLKRRPILVASTTTPGSGPSGCSISGADSTIRSLPYL
LDGCHPSMRKLIIAIVLIVPSIILAGVLLTVAIPGLSSVISPACMGALCIC
VMALCIDVLKRRPILVASTTTPGSGPSGCSISGADSTIRSLPYLPDDEHP
OSMRKLIIAIVLIVPSIILAGVLLTVAIPGLSSVISPACMGALGCVMLALG
IDVLKRRPILVASTTTPGSGPSGCSISGADSTIRSLPYLPDDEHP

CDS

VSHELNKMSLPDGHGLEBETKQIIRVRSLSKAVPEFLDIRIFEEFEFFLSAR
KRILDIATTLVERKILTEODERNLNKARSYKDSIFKKIIDNEFKLAKMFILSKS
ICQFTIIFENHENGVAKSLILKRAVLLKVEIYRSLQKSYRDKMSSAKCIIAGNFP
SLIEDNKTTIMKEHAEMLESLSYKRAVFLASDBNVVDTPSDRAKMLSCITPCDASE
ISRDQWQKAHLKHQESLYTQARDLTDQSSKENQKLEKAQETISSERVKKFEI
ERYQERIRAIQKLYPNILEREETIQEFTVTPVQGTASDITDILGRIEVSREDN
QUNQESCVKLRSHEVMSMEVKEOYPPKKERODQSGSLERFETHIELEIVSKQYS
KILSYFKYNNKKEVOYAKFRKLKESDLEGLIAOTESASLITQELPILATRGLE
KAFKSLICCALASKAKPYFEEDPREDDSDTQIRALTLLQEKASLEBEIRFSNLE
NDIAEERRLKESKOTFERAGAGVLEIVESTYTDLSLTNNWEGTPESEKVFYSMTL
NYYNEERKRAKTRILVEMTQYRDKMALEAMOFNEBALLQEEELISIQAPSE"

7150..8058

/gene="hb2"

7150..8058

/gene="hb2"

/note="hypothetical"

/codon_start=1

/transl_table=11

/product="Hb2 protein"

/protein_id="CAB37077.1"

/db_xref="GI:4455893"

/translation="MQVFLSPQLPPPPRRSLGYSCKLNSITTLVLVGLVLLITCIP
GLTAGISFGAGLQFSAAGVLVISGLFLVRRVPTVASEEIPRGVSTPSEPALE
KQKQEPETKILDRLEPRLDQDTYIOEFACIEKLDKPYEDRGILLTBAKEXLRVFD
VVEKDMSEFLDIQVLYNEEAYVEHCQDPLENIAYEIFSSQELRDQYCAQVGYLPS
GDARADRLKRSYKEVMDRPMRYTWKSGMEASVMDHSYGVARELFKKAQVLEESVYKI
LPSYRDAPYECCAKIQRDGRFKML"

8185..9492

/gene="hb3"

8185..9492

/gene="hb3"

/note="hypothetical"

/codon_start=1

/transl_table=11

/product="Hb3 protein"

/protein_id="CAB37078.1"

/db_xref="GI:4455894"

/translation="MMLDRVADKFLIREKEEKMERHELPHATVYRKASGAAVAKAA
FEKERENMORVKVYDEKMLSGLAFFRNQESRARERLEIOTITLPEVSEPRVIR
OPRRKYNLENIYADIKKYHHCYRDEQHTWKVEKNKAEYRENGEVLISAEVSEILO
RLBDCLETSKSLTKAESEVFEKFDATKELKKNVLDVYNNRLLEILCEAAEENIFRIE
ELEMTRMVELLFLMKNTEFKASLOYNCKEKLAKVFEQCKESPTVRSOEBLERLN
ODQLOTYVTCOERLQFSDLESKVRTRDRLREOMKLFVQCKINFTNEBLLWVGALE
TQARLDLVATVPMEFYLYOHNIKREKVSOMMAKTERYREIIOFQVGMKEDLLAED
TLKEEDVYLLRDMWLRLDERKKNRQRLCNKIAAQGVKGF"

9805..10698

/gene="hb4"

9805..10698

/gene="hb4"

/note="hypothetical"

/codon_start=1

/transl_table=11

/product="Hb4 protein"

/protein_id="CAB37079.1"

/db_xref="GI:4455895"

/translation="MLLISGALFLTLGIPGSAATSPGLGIGLSALGCVLMTSGLLC
LVKKRPIPVYRPEEIPQCVSLASEPAPLQAAQKTLAQUPKRELDQDTIOQVPAFLR
KLKDSKYESRSFLNDARKKEIRYDFVEQTLSTIFELROI VAOEGDNLNLINGGRSL
MMTAESESLDLPHVSKRLGYLPGSDVGRGGLKSAKEIYARMLNHCIELKHAVAADR
NSYAAAEKAFKALGALLESYVRSLTQSYRDXFLSEERAKIPMNGHITWLRODAGSC
AEKIKGMPPRVNGLGKOSFG"

10776..12119

/gene="hb5"

10776..12119

/gene="hb5"

/note="hypothetical"

/codon_start=1

/transl_table=11

/product="Hb5 protein"

/protein_id="CAB37080.1"

/db_xref="GI:4455896"

/translation="MDEIFHLEYKNTFLKRLHRKCALAKTTFEKKRSKKNLOAVEE
ANARRLKVVYRDMYDOEFQAGERLEKRLHALYEVSVSIRENNKIQETRSMLKAYEAI

gene

8185..9492

/gene="hb3"

8185..9492

/gene="hb3"

/note="hypothetical"

/codon_start=1

/transl_table=11

/product="Hb3 protein"

/protein_id="CAB37078.1"

/db_xref="GI:4455894"

/translation="MMLDRVADKFLIREKEEKMERHELPHATVYRKASGAAVAKAA
FEKERENMORVKVYDEKMLSGLAFFRNQESRARERLEIOTITLPEVSEPRVIR
OPRRKYNLENIYADIKKYHHCYRDEQHTWKVEKNKAEYRENGEVLISAEVSEILO
RLBDCLETSKSLTKAESEVFEKFDATKELKKNVLDVYNNRLLEILCEAAEENIFRIE
ELEMTRMVELLFLMKNTEFKASLOYNCKEKLAKVFEQCKESPTVRSOEBLERLN
ODQLOTYVTCOERLQFSDLESKVRTRDRLREOMKLFVQCKINFTNEBLLWVGALE
TQARLDLVATVPMEFYLYOHNIKREKVSOMMAKTERYREIIOFQVGMKEDLLAED
TLKEEDVYLLRDMWLRLDERKKNRQRLCNKIAAQGVKGF"

9805..10698

/gene="hb4"

9805..10698

/gene="hb4"

/note="hypothetical"

/codon_start=1

/transl_table=11

/product="Hb4 protein"

/protein_id="CAB37079.1"

/db_xref="GI:4455895"

/translation="MLLISGALFLTLGIPGSAATSPGLGIGLSALGCVLMTSGLLC
LVKKRPIPVYRPEEIPQCVSLASEPAPLQAAQKTLAQUPKRELDQDTIOQVPAFLR
KLKDSKYESRSFLNDARKKEIRYDFVEQTLSTIFELROI VAOEGDNLNLINGGRSL
MMTAESESLDLPHVSKRLGYLPGSDVGRGGLKSAKEIYARMLNHCIELKHAVAADR
NSYAAAEKAFKALGALLESYVRSLTQSYRDXFLSEERAKIPMNGHITWLRODAGSC
AEKIKGMPPRVNGLGKOSFG"

10776..12119

/gene="hb5"

10776..12119

/gene="hb5"

/note="hypothetical"

/codon_start=1

/transl_table=11

/product="Hb5 protein"

/protein_id="CAB37080.1"

/db_xref="GI:4455896"

/translation="MDEIFHLEYKNTFLKRLHRKCALAKTTFEKKRSKKNLOAVEE
ANARRLKVVYRDMYDOEFQAGERLEKRLHALYEVSVSIRENNKIQETRSMLKAYEAI

CDS

[illegible]


```
Align seg 1/1 to reverse of: CHOMP54 from: 1 to: 6030

12 CysAlaIaIleLeuSerSerThrAlaValLeuPheGlyGlnAspProLe 28
   |||.....:|||||.....:|||||.....:
2685 TGTTCACAGTGTGTTGTCGCAAGTCTGTAATAATATAGCCCTCGATAG 2636
   |||.....:|||||.....:|||||.....:
28 u.....cTgCtuThrAlaLeuLeuThrLysAsnProAsnHisValaIC 43
   |||.....:|||||.....:|||||.....:
2635 CTTTGACGGAAGTACTAACACAGGACCTATCTCTTAATAATAGACTA 2586
   |||.....:|||||.....:|||||.....:
43 ys.....ThrPheheGluAspCysThrMeGluSerLeu... 54
   |||.....:|||||.....:|||||.....:
2585 CTGCAATAGACTATACCTCTGACAGAGATATATACCTGCAAAACCTGCG 2536
   |||.....:|||||.....:|||||.....:
55 ..PheProAlaLeuCysAlaHisAlaSerGlnAsp.....AspProLe 68
   |||.....:|||||.....:|||||.....:
2535 GATTGCGAGCTTTATACGAGAGGTGTTTCTGACACTAGCAAGATCTTT 2486
   |||.....:|||||.....:|||||.....:
68 uTyrValLeuGlyAsnSerTyrCysTyrPheValSerLysLeuHisLiet 85
   |||.....:|||||.....:|||||.....:
2485 AAGCTTTCCCGGTAGGGGTACTCA.....CTTCTTTTAAATATTA 2442
   |||.....:|||||.....:|||||.....:
85 hrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerLiegln 101
   |||.....:|||||.....:|||||.....:
2441 AGCTACGTGCGCAAGCC.....GCACGACTTTCGTTCACA 2407
   |||.....:|||||.....:|||||.....:
102 AsnPheArGpHeLeuSerPheThrAspCysSerSerLysGlu.....Se 116
   |||.....:|||||.....:|||||.....:
2406 ACTGTATTAATAATCTGTCTGCTAACAGGATTTTCGAGTCTTACTTCTT 2357
   |||.....:|||||.....:|||||.....:
116 rSerProSerLietLe.....HisGlnL 124
   |||.....:|||||.....:|||||.....:
2356 GGCCCATCATCGGTGTAATCACAACCCCTCAGAAAAGCTCAGTTAAT 2307
   |||.....:|||||.....:|||||.....:
124 ysAsnGlyGlnLeuSerLeuArGAsnAsnGlySerMetSerPheCysAr 140
   |||.....:|||||.....:|||||.....:
2306 CTGCAAGGAGATCTACATTTGATTAACAATGACATTTTATTTAACAA 2257
   |||.....:|||||.....:|||||.....:
141 AsnHisAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSerLe 157
   |||.....:|||||.....:|||||.....:
2256 GATTACTGTGGGAAATGGCGGAGCCATTTCTACAGAAATCTTTCTTT 2207
   |||.....:|||||.....:|||||.....:
157 uGlnHisAsnTyrLeuPheThrAlaPheGluGlnAsnSerSerLysGly. 173
   |||.....:|||||.....:|||||.....:
2206 GAAAACAGCACGGGATGATTTCTTTGAAAGGCAATTAATCGACGCCAA 2157
   |||.....:|||||.....:|||||.....:
174 .....AsnGlyGlyAlaIleGlnAlaGln...ThrPheSerLeuSer 186
   |||.....:|||||.....:|||||.....:
2156 CACGCAAAAAGGTGGGCTATTGTCTACTGTAAGTATGATATTTACA 2107
   |||.....:|||||.....:|||||.....:
187 ArGAsnValSerProIleSerPheAlaArGAsnArGAlaAspLeuAsnG 203
   |||.....:|||||.....:|||||.....:
2106 AATAATACGGCTCTACCTCTCTCGAACAATATTGCTGAAGCTGCAGG 2057
   |||.....:|||||.....:|||||.....:
203 yGlyAlaIle...CysCysSerAsnLeuIleCysSerGlyAsnValAsn 219
   |||.....:|||||.....:|||||.....:
2056 TGGAGCTTAATAATACAGCAAGAACTGACAAATTACAGGAATACGCT. 2008
   |||.....:|||||.....:|||||.....:
219 rOleuPhePheThrGlyAsnSerAlaThr.....AsnGlyGly 231
   |||.....:|||||.....:|||||.....:
2007 ..CTTGATTTTTCGAAATAAGTGTGACAGCACGCCACGAAATGACAGA 1960
   |||.....:|||||.....:|||||.....:
232 AlaIleCysCysIleSerAspLeuAsnThrSerGlyLysGlySerLeuSe 248
   |||.....:|||||.....:|||||.....:
1959 GCTCTT.....TCTGAGATGCCGATGTGTAC 1934
   |||.....:|||||.....:|||||.....:
248 rLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysGluL 265
   |||.....:|||||.....:|||||.....:
1933 CATATCTGGAGATCAGAGTGTACTTTCTCAGAAACCAAGCTGAGCTA 1884
   |||.....:|||||.....:|||||.....:
265 ysglyGlyAlaIleTyrAlaLysHisMetValLeuArGtyr.....Asn 279
   |||.....:|||||.....:|||||.....:
1883 ATGCCGAGACCATTTATCTCTAAGAGCTTACACTGCTTCGCGGCGCGGG 1834
   |||.....:|||||.....:|||||.....:

280 GLyProValSerPheIleAsnAsn.....SerAlaLysI 291
   |||.....:|||||.....:|||||.....:
1833 GGGGATACCTCTTTCTTACAAATATAGTCCAGATACCACTGACAGGTAA 1784
   |||.....:|||||.....:|||||.....:
291 eGlyValaIleAlaIleGlnSerGlyGlySerLeuSerLietLeuAlaG 308
   |||.....:|||||.....:|||||.....:
1783 TCGTGAACCTTTCTTATCTGCGACCTGCGAGAGCTGATCTTTCAGCAG 1734
   |||.....:|||||.....:|||||.....:
308 LyGluGlySerValLeuPheGlnAsnAsnSer...GlnArGThrSerAsp 323
   |||.....:|||||.....:|||||.....:
1733 AAGCAGCGGACATTAACCTTCAATGGAAATGCCATTTGCAATACACCA 1684
   |||.....:|||||.....:|||||.....:
324 GlnGlyLeuValArGAsnAlaIleTyrLeuGluLysAspAlaIleLeuSe 340
   |||.....:|||||.....:|||||.....:
1683 CAACCTACAAAAGAAATTCATTATGACATAGACTACTGCAAGAGATCAC 1634
   |||.....:|||||.....:|||||.....:
340 rSerLeuAlaIleArGAsnGly...AspIleLeuPheAspProLiev 356
   |||.....:|||||.....:|||||.....:
1633 GAATTTACGTCAATATCTGGCATAGCATCTTTTCTAGCATCCGATTA 1584
   |||.....:|||||.....:|||||.....:
356 aLglnGlySerSerSerLysGluSerProLeuProSerSerLieglnAla 372
   |||.....:|||||.....:|||||.....:
1583 CTGCTAATAGCCGTGCGGATTCATACAGATCTTAATCTCAATAGAGCT 1534
   |||.....:|||||.....:|||||.....:
373 SerValThrSerProThrProAlaThrAlaSerProLeuValIleGlnTh 389
   |||.....:|||||.....:|||||.....:
1533 GATCGAGCTAAT.....AGTAC 1517
   |||.....:|||||.....:|||||.....:
389 rSerAlaAsnArGSerValIlePheSerSerGluArGLeuSerGluG 406
   |||.....:|||||.....:|||||.....:
1516 AGATTTATAGTGGCTGATGTTTCTGTGGAAGCTCTCTCAAGATG 1467
   |||.....:|||||.....:|||||.....:
406 Lu...LysThrProAspAsnLeuThrSerGlnLeuGlnGlnProIleGlu 421
   |||.....:|||||.....:|||||.....:
1466 AAGCAAAAGTTGCAAGACACTCTTACCTGCAAGACACCTGTAAT 1417
   |||.....:|||||.....:|||||.....:
422 LeuLysSerGlyArGLeuValLeuLysAspArGAlaValLeuSerAlaPr 438
   |||.....:|||||.....:|||||.....:
1416 CTACTGCGAGAAATTTAGTCTTAACGTGTGTCTCTCGATAGCAA 1367
   |||.....:|||||.....:|||||.....:
438 cSerLeuSerGlnAspProGlnAlaLeuLeuIleMerGluAlaGlyThrS 455
   |||.....:|||||.....:|||||.....:
1366 AGCCTTACTGACAGCCGGGTCTCTGTTTATATGATCGGCGACAA 1317
   |||.....:|||||.....:|||||.....:
455 eLeuLysThrSerSer...AspLeuLysLeuAlaThrLeuSerIlePro 470
   |||.....:|||||.....:|||||.....:
1316 CGTTAAAGCAAGTACAGAGAGGTCATTAACAGGCTTTCCATTCCT 1267
   |||.....:|||||.....:|||||.....:
471 LeuHisSerLeuAspThrGluLysSerValThrIleHisAlaPro.... 485
   |||.....:|||||.....:|||||.....:
1266 GTAGACTCTTTAGGCGAGCGGTAAAGATTTGTAATGCTGCTTCGACG 1217
   |||.....:|||||.....:|||||.....:
486 .....AsnLeuSerIleGlnLysIlePheLeuSerAsnSerG 498
   |||.....:|||||.....:|||||.....:
1216 AAGTAAATAATATACCCCTTACTGCTCGGATTTCTTTTGGATACCAAG 1167
   |||.....:|||||.....:|||||.....:
498 LyAspGluAsnPheTyrGluAsnValGluLeuLeuSerLysGluGlnAsn 514
   |||.....:|||||.....:|||||.....:
1166 GG.....AATGCTTATGAAATACAGAC...TTAGGAAAATCAGAC 1126
   |||.....:|||||.....:|||||.....:
515 .....AsnIleProLeuLeuThrLeuSerLysGluGlnSerHi 527
   |||.....:|||||.....:|||||.....:
1125 TTTTCATTGTGACAGCTCTGTGCTGGTACTGCAACAACATACAGATGT 1076
   |||.....:|||||.....:|||||.....:
527 sLeuHisLeuProAspGlyAsnLeuSerSerHisPheGlyTyrGlnGly 544
   |||.....:|||||.....:|||||.....:
1075 TCCAGCGGTTCTTACAGTACGAACTCTTACGACTATATGGATATGAAGTA 1026
   |||.....:|||||.....:|||||.....:
544 sPTrPThrPheSerTrpLysAsp.....SerAspGluGlyHis 556
   |||.....:|||||.....:|||||.....:
1025 CTTCGGGAAATGACTTCGCTGATGATACCGCAAGACACTCCAAAGACTAAG 976
   |||.....:|||||.....:|||||.....:
```

994 ..CTTGATTTCGAAATAGTGCACGACCGCAGCAATATGAGCA 1041
232 AlaIleCysCysIleSerAspLeuSerGluLysGlySerLeu 248
1042 GCTCTTTCTGGAGATGCCGATGTC 1067
248 rLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysGlu 265
1068 CATATCTGGAAATCAGAGTGTACTTCTCAGGAAACCAAGCTGAGCTTA 1117
265 yGlyGlyAlaIleTyrAlaLysHisMetValLeuArgTyrAsn 279
1118 ATGGCGGACCATTTATGTAGAACCTTACACTGCGCTCCGGGGGGGG 1167
280 GlyProValSerPheIleAsnAsnSerAlaLysIle 291
1168 GGGGATATCTCTTCTTCAACAATATAGTCAAGGTACCACTGACAGTAA 1217
291 eGlyGlyAlaIleAlaIleGlnSerGlyLysSerLeuSerIleLeuAlaG 308
1218 TGGTGGAGCCATTCTATCTAGTGCAGAGTGAAGTGTAGTCTTTCAGCAG 1267
308 LysGluLysValLeuPheGlnAsnAsnSer...GlnArgThrSerAsp 323
1268 AAGCAGGACATTTACTCTCATTCGCAATGCCATCTTCACACTACACA 1317
324 GlnGlyLeuValArgAsnAlaIleTyrLeuGluLysAspAlaIleLeu 340
1318 CAACACTACAAAGAAATCTATGTACATAGATCTACTGCAAAAGATCAC 1367
340 rSerLeuGluAlaArgAsnGly...AspIleLeuPhePheAspProIle 356
1368 GAATTTACTGCATATCTGGGCATAGCTTTCTTCTACGATCCGATTA 1417
356 aGlnGlnLysSerSerLysGlnSerProLeuProSerSerLeuGlnAla 372
1418 CTGCATATACGGCTCGCGATCTTACAGATCTTAAATCTCAATAAGCTT 1467
373 SerValThrSerProThrProAlaThrAlaSerProLeuValIleGlnThr 389
1468 GATGACGATAATAGTAC 1484
389 rSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerGluGlu 406
1485 AGATTATAGTGGTCGATTGTTTCTGTGTAAGCAAGCTCTGTAAGAG 1534
406 Lu...LysThrProAspAsnLeuThrSerGlnLeuGlnProIleGlu 421
1535 AAGCAAAAGTTGCAGACACCTCTACCTACGCTGACAGAGCTCTACT 1584
422 LeuLysSerGlyArgLeuValLeuLysAspArgAlaValLeuSerAla 438
1585 CTAACTGCAGAAATTTACTACTTAACGTGTGTCACCTCGATACGAA 1634
438 oSerLeuSerGlnAspProGlnAlaLeuLeuIleMetGluAlaGlyHis 455
1635 AGGCTTTACTAGACCGCGGTTCTCTGTATTATGTAGTGGGACAAA 1684
455 rLeuLysThrSerSer...AspLeuLysLeuAlaThrLeuSerIlePro 470
1685 CATTAAACCAAGTACAGAGAGCTACTTTACAGGCTTTCCATTCTCT 1734
471 LeuHisSerLeuAspThrGluLysSerValThrIleHisAlaPro..... 485
1735 GTAGACTCTTAGCGAGGTAAAGATTGAATTGCTTTCGACAGC 1784
486AsnLeuSerIleGlnLysIlePheLeuSerAsnSerG 498
1785 AAGTAAATATGATGCTTAGTGTCCGATTTCTTTGGATAACCAAG 1834
498 LysAspGluAsnPheTyrGluAsnValGluLeuLeuSerLysGlnGln 514
1835 GC.....AATGCTTATGAATAATCACGAC...TTACGAAAAACCTCAAGAC 1875

515AsnIleProLeuLeuThrLeuSerLysGlnGlnSerIle 527
1876 TTTCATTTGTGCAGCTCTGCTGCTGGTACTGCACAACATACAGATGT 1925
527 sLeuHisLeuProAspGlyAsnLeuSerSerHisPheGlyTyrGlnGly 544
1926 TCCACCGGTCTCTACAGTAAAGTCTTACGACATATGCTATACAGTAA 1975
544 sPThrThrPheSerThrLysAspSerAspGluGlyHis 556
1976 CTGGCGAATGACTTGGTTCATGATACCGCAGACACTCCAAAGCTTAAG 2025
557 SerLeuIleAlaAsnThrPheProLysAsnTyrValProHisProGluArg 573
2026 ACAGGACATTTAGTGTGACCAATAGAGCTACCTCGAATCTCGAGCG 2075
573 gGlnSerThrLeuValAlaAsnThrLeuThrAsnThrTyrSerAspMetG 590
2076 TCAAGACCTTTAGTCTCAATATACCTTTGGGAGCTTTTCACAGATCC 2125
590 LysAlaValGlnSerMetIleAsnThrIleAlaHisGlyAlaTyrLeu 606
2126 AAGCATTTCAAGCTGTCTACAGACAGTCTTTCAGCTTTGTTACAT 2175
607 PheGlyThrTyrGlySerAlaValSerAsnLeuPheTyrAlaHisAspSe 623
2176 CGAGCTCTTGGCTGCGCGAGTCCCAAT...TTCTTAGTAATAAGATA 2222
623 rSerGlyLysProIleAspAsnThrPheHisArgSerLeuGlyTyrLeuP 640
2223 GAAAGCGGAAAA...CCCAATACCGTCATATAATCTGTGGATATGCTA 2269
640 heGlyIleSerThrHisSerLeuAspAspHisSerPheCysLeuAla 656
2270 TCGCAGGTCCAGCGCAACTGTTCTGAAACTTAATTAAGCTTTGCCCTT 2319
657 GlyGlnLeuLeuGlyLysSerSerAspSerPheIleThrSerThrGluThr 673
2320 TCCCAACTCTTGTGATGAGATAAAGATTCTTAGCGCTAAATAATCATAC 2369
673 rThrSerTyrIleAlaThrValGlnAlaGlnLeuAlaThrSerLeu... 688
2370 TGATATCTATGACAGAGCTCTGTATATCCAAACATTCAGCAATGATG 2419
688 688
2420 GGTTCATAGCTTCTCTTACATTAACCTCTGCTGAGATCATATAA 2469
689 ...MetLysIleSerAlaGlnAlaCysTyrAsnGlnSerIleHisGluLe 704
2470 CCCCCTGTTTAAAGAGGACCTGCTATAGCCAGCTCATATATGATCT 2519
704 uLysThrLysTyrArgSerPheSerLysGlyGlyPheGlySerThrHis 721
2520 GAAGCAAACTATACCTCGTAT...CTGAGGTGAAGAGCTTTGGGGCA 2566
721 rValAlaValSerGlyLysAlaSerIleProIleValSerAsn 737
2567 ATATCTCTTTTAACTATGATGTGGAGCTTCTCTCATTTTATCTTAA 2616
738 GlySerGlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGlyPh 754
2617 TACGTGATGTTTGTGATACCTATGCTCATACATCAAGATGTGAC 2666
754 eSerGlyThrGlnAspGlyPheGluGlnSerSerGlyGluLeuArgSerP 771
2667 CTATATACGTAGAGACACTTCTCGAGAAAGTACAGAAAGAGATCTT 2716
771 heSerAlaSerSerPheArgAsnIleSerLeuProIleGlyIleThrPhe 787
2717 TTGATGACAGCAACTCTTCAATTATCTTTCCTATATAGGGGTAAGTTT 2766

```
633 ACGGGAAGTCTTTTTCGACAACTTCGCGATTCGACCTGCAA 682
206 IecysCysSerAsnLeuIleCysSerGlyAsnValAsnProLeuPhePhe 222
    |||
683 TTTTACAGAAAGCTCGTGACTATTCTTAATAAGTAAAGTTCTT 732
223 ThrCysAsnSerAlaThr.....As 229
    |||
733 ATTGACAATAAGTACACAGAGAGCTCTCAACACAGGGGATATGTC 782
229 nglyGlyAlaIleCysCysIleSerAspLeuAsnThrSerGlyGly 246
    |||
783 AGACCTGCTATCTGTCT.....TATAAACTACTACAGATCTA 823
246 erLeuSerLeuAlaCysAsnGlnGluThrPheAlaSerAsnSerAla 262
    |||
824 AGGTCACTCCTCAGTCAATCAGATGTTACTTTCAGACACATACATCG 873
263 LysGlyLysGlyAlaIleTyrAlaLysHisMetValLeuArgTyrAs 279
    |||
874 ACACACCGGAGAGAGCTATCTATGTGAAAAGCTCGACTGGCTCGG 923
279 nglyProValSerPheIleAsnAsnSer.....AlaLysI 291
    |||
924 AGCACTTACCTTCTCAGTAGAAATCTCTCAATGAGTACACCTCTA 973
291 IeglyGlyAlaIleAlaIleGlnSerGlyLysSerLeuSerIleLeuAla 307
    |||
974 AAGTGAAGCATACGATACGAGATGAGTGGGATTTGATTCGCCGC 1023
308 GlyGluGlySerValLeuPheGlnAsnAsnSerGlnArgThrSerAspG 324
    |||
1024 GATAGTGTCACATTTCTTTTAGGGAATACAGTCACTCTACTCTCC 1073
324 nglyLeuValArgAsnAlaIleTyrIleGlyLysAspAlaIleLeuSer 341
    |||
1074 TGGGACCAATAGAGATGATTCGATTCAGAACCGCAAGATGACAG 1123
341 erLeuGluAlaArgAsnGly...AspIleLeuPheAspProIleVal 356
    |||
1124 CTTTGCTTCTGCTGCTGTAGACCATCTACTTCTATGATCCCATACT 1173
357 GlnGluSerSer.....LysGluSerP 365
    |||
1174 ACAGATCTTCCACACAGTACAGATGCTTAAAGTAAATGAGACTCC 1223
365 OleuProSerSerLeuGlnAlaSerValThrSerProThrProAlaThrA 382
    |||
1224 GGCAGATTCGACTACATATACA..... 1248
382 IaserProLeuValIleGlnThrSerAlaAsnArgSerValIlePheSer 398
    |||
1249 .....GGGAACATCATCTTCA 1266
399 SerGluArgLeuSerGluGluLysThrProAsp...AsnLeuThrSe 414
    |||
1267 GGAGAAAGTATACAGAGACAGCGCGAGATTCCTAAATCTTACTTC 1316
414 rGlnLeuGlnProIleGluLeuLysSerGlyArgLeuValLeuLysA 431
    |||
1317 GAAGCTACTACAGCTGTAACTCTTTCAGAGGACTACTATCTTAAAC 1366
431 sPArgAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeu 447
    |||
1367 ATGAGAGTACTGTCAGACTCAGCATTCCTACACAGCGCATTCCTCGT 1416
448 LeuIleMetCysAlaGlyThrSerLeuLysThrSerSerAspLeuLysI 464
    |||
1417 CTGCAATGAGCTAGACTACTCTAGAACCTGATCTACTAGCACCAT 1466
464 uAlaThrLeuSerIleProLeuHisSerLeuAspThrGluLysSerValT 481
    |||
1467 AACCAATTTGCTATTAACTAGCTTCTATAGACCGCTGCAAAAGACAA 1516
481 hrIleHis.....AlaProAsnLeuSerIleGlnLysIlePhe 493
    |||
1517 AATAGAACCAAGCTACGTCAAAAATCTGACTTATCTGGAACCATC 1566
494 LeuSerAsnSerGlyAspGluAsnPheTyrGluAsnValGluLeuLeuSe 510
    |||
1567 ACTTTATTGGACCCGAGCGGACAGCTTTTATGAAATCATAGTTAAGAA 1616
510 rLysGluGlnAsnAsnIleProLeuLeuThrLeuSer...LysGluGlnS 526
    |||
1617 TCCCACTCTTACAGACTCTTACAGCTCAAAAGCTTCGAACTGAACAA 1666
526 erHisLeuHisLeuProAspGlyAsnLeuSer.....HisPheGly 540
    |||
1667 GCACCGCAGTACTCCAGATCTATATGAGTGAAGAAATTCATTACGGC 1716
541 TyrGlnGlyAspThr...ThrPheSerTrpLysAspSerAspGlyLysI 556
    |||
1717 TATCAGGCACTTCGGGCCCATTTGTTGC.....GGCAGACGGCC 1757
556 sSerLeuIleAla.....AsnTrpThrProLysAsnTyrValProHisP 571
    |||
1758 TTTCACACTCTCAACCTTCACATGACATAAAGCTGCTATATTCCTAATC 1807
571 rGluArgGlnSerThrLeuValAlaAspThrLeuTrpAsnThrTyrSer 587
    |||
1808 CCGAGCTATCGGCTCTTATGCTCCTATATAGCTTATGGAATGCAATTAA 1857
588 AspMetGlnAlaValGlnSerMetIleAsnThrIleAlaHisGly..... 602
    |||
1858 GATATTAGCTCTCTCCATTAATTCATGACACGCAACAGACGTTGCCA 1907
603 .GlyAlaTyrLeuPheGlyThrTrpGlySerAlaValSerAsnLeuPheT 619
    |||
1908 GCGAGACCTGCTTTT.....TGCTCTCGATGATATCAACTCTTC. 1950
619 rAlaHisAspSerSerGlyLysProIleAspAsnThrPheHisIleAspSer 635
    |||
1951 .....CATAGATAGTACAAAACACGACGCGGTTTCCGCCATTGAGT 1995
636 LeuGlyThrLeuPheGlyIleSerThrHisSerLeuAspHisSerPhe 652
    |||
1996 GCGCGTTATGTCATACAGACGAAACCTACATCTTGTACATTAAGATTCT 2045
652 eCysLeuAlaIleGlnLeuLeuGlyLysSerSerAspSerPheIleT 669
    |||
2046 TATCTCTGCAATTTCTGACGCTTTCGAAAGATAGAGACTACTTGTAG 2095
669 hrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaGlnLeuAla 685
    |||
2096 CTAGAGTCAAGGTACAGTCTACGAGGAACTCTATTAACGACACAC 2145
686 ThrSerLeuMetLysIle..... 691
    |||
2146 GAACCAATATCTCTCTCTGCTCAAACTACGCGCTTGTCTGTGCTTA 2195
692 .....SerAlaGlnAlaCysTyrAsnG 699
    |||
2196 TGTTCCTACAGAGATTCCTGCTCTTTCAGGAAACCTTACACACCC 2245
699 IuSerIleHisGluLeuLysThrLysTyrArgSerPheSerLysGluGly 715
    |||
2246 ATACGGATACGATCTGAAACCAAGTATACACATATCTTACTGTTAA 2295
716 PheGlySerThrPheHisSerValAlaValSerGlyGluValCysAlaSerI 732
    |||
2296 ...GGAAGCTGGGGAATGATGATTCGCTTGTGAATTCGGTGAAGAGC 2342
732 eProIleValSerAsnGlySerGlyLeuPheSerSerPheSerIlePheS 749
    |||
2343 TCCGATTTCCTTAGATGAAGATGCTCTATTTGAGCAGTACATCCCTTCA 2392
```

```
12 CysAlaAlaIleLeuSerSerThrAlaValLeuPheGlyGlnAspProLe 28
|||||
7593 TGTCCTACTGTTTTCGTCACAGCTGCTGAAAATATAGGCCCTCTGATAG 7544
28 u.....GlycylThrAlaLeuLeuThrLysAsnProAsnHisValAlC 43
|||||
7543 CTTCGACGAGTACTACACAGCACTATCTCTTAAATAATACGACTA 7494
43 ys.....ThrPheGluAspCysThrMetCJusSerLeu... 54
|||||
7493 CTGCAATGACTATCTCTGACAGAGATATACCTGCAAAACCTTGGG 7444
55 ...PheProAlaLeuCysAlaHisAlaSerGlnAsp.....AspProLe 68
|||||
7443 GATTCGCCAGCTTTAAAGAGAGGCTGTTTTCCTGACACTACGAGATCTTT 7394
68 uTrpValLeuLeuLysSerTrpCysTrpPheValSerLysLeuHisLier 85
|||||
7393 AAGCTTTCGCCGTAGGGGTACTCA.....CTTCTTTTAAATATTA 7350
85 hrAspProLysGluAlaLeuPheLysGluLysCysAspLeuSerLecJin 101
|||||
7349 AGCTAGTCTGGAAGGC.....CCAGCACTTTCCTGTATCA 7315
102 AsnPheArgPheLeuSerPheThrAspCysSerLysGlu.....Se 116
|||||
7314 ACTGATAAATCTGTCCCTAACAGACTTTTCCACTCTTCTCTTACG 7265
116 rSerProSerLecJin.....HisGlnL 124
|||||
7264 GGCCCATCATCGGTATACACACCCTCTCAGAAAAGGTGACGTTAAT 7215
124 ysAsnGlyGlnLeuSerLeuArgAsnGlySerMetSerPheCysArg 140
|||||
7214 CTGGAGGGGACTTACATTTGATTAACAATGAACTATTTTATTTAAAGAA 7165
141 AsnHisAlaGlySerGlyGlyAlaIleSerAlaAspAlaPheSerLe 157
|||||
7164 CATTAAGTGTGAGAAATCCCGGAGCCATTTCTACCAACAATCTTCTTT 7115
157 uGlnHisAsnTrpLeuPheThrAlaPheGluLysSerSerLysGly 173
|||||
7114 GAAAAACGACGAGGATGATTTCTTTTGAAGGGAATTAATCGACGCCAA 7065
174 .....AsnGlyAlaIleGlnAlaGln...ThrPheSerLeuSer 186
|||||
7064 CAGGAAAAAAGGTGGGCTATTTGCTCTACTGCTACTGATATATTTACA 7015
187 ArgAsnValSerProLysPheAlaArgAsnArgAlaAspLeuAsnG 203
|||||
7014 AATATACGCTCTACCTCTCTCTCGAACATATTCCTGAACTGCCAG 6965
203 yGlyAlaIle...CysCysSerAsnLeuIleCysSerGlyAsnValAsn 219
|||||
6964 TGGACTTAAATAGCACAGAACTGTACATTAACGGAATACGCTT 6916
219 rLeuPhePheThrGlyAsnSerAlaThr.....AsnGlyGly 231
|||||
6915 ..CTTGTAATTTCTGAATAATAGGTGACACGACCGCAGAAATCGACGA 6868
232 AlaIleCysCysIleSerAspLeuAsnThrSerGlyLysGlySerLe 248
|||||
6867 CCWCTT.....TCTGACATCCCGAGTGTAC 6842
248 rLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysG 265
|||||
6841 CATATCTGGATCGAGCTGTACTTCTCAGAAACCAACAGCTGTAGCTA 6792
265 yScjGlyAlaIleTrpAlaLysHisMetValLeuArgTyr.....Asn 279
|||||
6791 ATGGCGGAGCAATTTATCTAAGAAAGCTTACACTGCC..TTCGGCGGGGG 6743
280 GlyProValSerPheIleAsnAsn.....SerAlaLysI 291
|||||
6742 GGGGCTATCTCTTTTCAACAATATAGTCCAGGTACACTGCAAGTAA 6693
291 eGlyAlaIleAlaIleGlnSerGlyCysLeuSerLecJinLeuAlaG 308
|||||
6692 TGGTGAGACCATTTCTATATGCGAGCTGAGAGTGTAGTCTTTCAGAG 6643
308 LysGlySerValLeuPheGlnAsnAsnSer...GlnArgThrSerAsp 323
|||||
6642 AAGCAGCGGACATTAATCTTCAATGGGAATGCCATTTGTCACACTACA 6593
324 GlnGlyLeuValArgAsnAlaIleTrpLeuGluLysAspAlaIleLe 340
|||||
6592 CAACCTACAAAAAGAAATCTATTGACATAGATCTACTGCAAGATCTAC 6543
340 rSerLeuGluAlaArgAsnGly...AspIleLeuPhePheAspProL 356
|||||
6542 GAATTTACGTCAATATCTGCGCATAGCATCTTTTCTACGATCCGATTA 6493
356 aGlnGlySerSerSerLysGluSerProLeuProSerSerLeuGlnAla 372
|||||
6492 CTGCTAATACGGCTGGGATTTCTACAGATCTTAAATCTCAATAGGCT 6443
373 SerValThrSerProThrProAlaThrAlaSerProLeuValIleG 389
|||||
6442 GATGAGGATAT.....AGTAC 6426
389 rSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerGluG 406
|||||
6425 AGATTATAGTGGTGATGTTTCTTTCGTGGAAGCTCTCTGAAGATG 6376
406 Lu...LysThrProAspAsnLeuThrSerGlnLeuGlnGlnProL 421
|||||
6375 AAGCAAAAGTTGACAGACACCTCTTACCGTGAACACCTGTACTACT 6326
422 LeuLysSerGlyArgLeuValLeuLysAspArgAlaValLeuSerAla 438
|||||
6325 CTACTCGAGGAATTTAGTACTTAACGTTGTGTCTACTCGATACGAA 6276
438 oSerLeuSerGlnAspProGlnAlaLeuLeuIleMetCJuaGlyThrS 455
|||||
6275 AGGCTTTACTCGACCGCGGGTCTCTCTATATATGATGCGGCGACAA 6226
455 eLeuLeuThrSerSer...AspLeuLysLeuAlaIleThrLeuSerL 470
|||||
6225 CTTTAAAGCAGATACAGAGGCTCACTTATACAGCTCTTCCATTCCT 6176
471 LeuHisSerLeuAspThrGluLysSerValThrIleHisAlaPro... 485
|||||
6175 GTAGACTTTTAGCGGAGGTAAGAAGTGTATGTGCTGCTTTCGACG 6126
486 .....AsnLeuSerLecJinLysIlePheLeuSerAsnSerG 498
|||||
6125 AAGTAAATATGACCCCTTACTGTCGATTCCTTTTGGATTAACCAAG 6076
498 LysAspGluAsnPheTrpGluAsnValGluLeuLeuSerLysGluGln 514
|||||
6075 GG.....AATGCTTATGAATAATCACGAC...TTAGAAAACATCAAGAC 6035
515 .....AsnIleProLeuLeuThrLeuSerLysGlnGlnSerH 527
|||||
6034 TTTTCATTTTGACAGCTCTGCTGCTGGCTGCTGCAACAATACAGATGT 5985
527 sLeuHisLeuProAspGlyAsnLeuSerSerHisPheGlyTyrGlnG 544
|||||
5984 TCCAGCGGCTCTCTACAGTACGCAACTCTACACACTATAGGATACAGTA 5935
544 sTrpThrPheSerTrpLysAsp.....SerAspGluGlyHis 556
|||||
5934 CTGGCGCATGACTTGGTGTATATACCGCAACACTCCCAACACTPAG 5885
557 SerLeuIleAlaAsnTrpThrProLysAsnTrpValProHisProGlu 573
|||||
```

```
222 AlaIleCysCysIleSerAspLeuAsnThrSerGluLysGlySerLeuSe 248
|||||
|||||
207047 GCTCTT.....TCTGAGATGCCGATGTTAC 207022
248 rLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysGluL 265
|||||
|||||
207021 CATATCTGGAGTACAGACTGTAACCTTCTCAGAGAAACCAAGCTGTACGA 206972
265 YSGLYGLVALAAlaIleTyrAlaLysHisMetValLeuArgTyr.....Asn 279
|||||
|||||
206971 ATCGCGAGACCATTTATCTCTAAGAAGCTTACAGCTGGCTTCCGCGGGGCG 206922
280 GlyProValSerPheIleAsnAsn.....SerAlaLysIle 291
|||||
|||||
206921 GGGGATATCTCTTCTTCTAACAATATAGTCCAGGTACCGAGGTAA 206872
291 eGlyGlyAlaIleAlaIleGlnSerGlyGlySerLeuSerIleLeuAlaG 308
|||||
|||||
206871 TCGTCGACCCATTCTCTATACGCGACGACGACGCTAGCTCTTCACGACG 206822
308 YLGLGLYSerValLeuPheGlnAsnAsnSer...GlnArgThrSerasp 323
|||||
|||||
206821 AAGCAGGAGACATTACCTTCAATGGAATGCCATTCTTGCAACTACACCA 206772
324 GlnGlyLeuValArgAsnAlaIleTyrLeuGluLysAspAlaIleLeuSe 340
|||||
|||||
206771 CAACACTAGAAAGAAATTTCTATTGACATAGATCTACTGCAAGAATACAC 206722
340 rSerLeuAlaIleArgAsnGly...AspIleLeuPheAspProIleVal 356
|||||
|||||
206721 GAATTTACGTCCAATATCTCGCATACATCTTTTCTACAGATCCGATTA 206672
356 aGlnGlnSerSerSerLysSerLysSerProLeuProSerSerLeuGlnAla 372
|||||
|||||
206671 CTGCTAATACCGCTCGGATTTCTACAGATCTTAAATCTCAATAGGCT 206622
373 SerValThrSerProThrProAlaThrAlaSerProLeuValIleGlnTh 389
|||||
|||||
206621 GATGCGAGTAAAT.....AGTAC 206605
389 rSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerGluGluG 406
|||||
|||||
206604 AGATTTATAGTGGCGATGCTTTTCTGCGTGAAGAGCTCTCTGAAAGATG 206555
406 Lu...LysThrProAspAsnLeuThrSerGlnLeuGlnAlaProIleGlu 421
|||||
|||||
206554 AAGCAAAAGTTGACAGACAACCTCACTTACGCTGAAGACGCTTAACT 206505
422 LeuLysSerGlyArgLeuValLeuLysAspArgAlaValLeuSerAlaPr 438
|||||
|||||
206504 CTAACCTGACGAAATTTAGTACTTAACGTGGTGCACCTCTCGATGACGA 206455
438 oSerLeuSerGlnAspProGlnAlaLeuLeuIleMetClnAlaGlyThrS 455
|||||
|||||
206454 AGGCTTTACTGACAGACCCCGGCTCTCTGTATATATGATCGCGGACAA 206405
455 eLeuLysThrSerSer...AspLeuLysLeuAlaThrLeuSerIlePro 470
|||||
|||||
206404 CCTTAAAGCAAGTACAGACAGAGCTCTTAAACAGCTTTTCCATTTCT 206355
471 LeuHisSerLeuAspThrGluLysSerValThrIleHisAlaPro.... 485
|||||
|||||
206354 GTAGACTCTTGAAGGAGGTAAAGAAAGTGTATGCTGCTTGCAGC 206305
486 .....AsnLeuSerIleGlnLysIlePheLeuSerAsnSerG 498
|||||
|||||
206304 AAGTAAAAATGTAGCCCTTACTGCTCGATTTCTTTTGATAAACCAAG 206255
498 lYAspGluAsnPheTyrGluAsnValGluLeuLeuSerLysGlnGluAsn 514
|||||
|||||
206254 GG.....AAIGCTTATGAAATATCAGCAC...TTAGCAAAAACCTACAGC 206214
```

```
515 .....AsnIleProLeuLeuThrLeuSerLysGlnGlnSerI 527
|||||
|||||
206213 TTTCATTTGACAGCTCTCTGCTGTACTGCAACACTACAGATGT 206164
527 sLeuHisLeuProAspGlyAsnLeuSerSerHisPheGlyTyrGlnGlyA 544
|||||
|||||
206163 TCCACCGGCTTCTACAGTACCAACTCTACGCACTACTATGGTATCAAGCTA 206114
544 sPTPrThrPheSerTrpLysAsp.....SerAspGlnGlyHis 556
|||||
|||||
206113 CTGGGGATGATCTGGGTATGATGATACCGCAAGCAGCTCAAGAGCTAAG 206064
557 SerLeuIleAlaAsnTrpThrProLysAsnTyrValProHisProGluArg 573
|||||
|||||
206063 ACAGGACATTAGCTGTGACCAATACAGCTACCTTCGGAATCTGTAGCG 206014
573 gGlnSerThrLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMetG 590
|||||
|||||
206013 TCAAGACCTTTAGTTCTCTATAGCCTTGGGATCTTTTTCAGACATCC 205964
590 lAlaValGlnSerMetIleAsnThrIleAlaHisGlyAlaTyrLeu 606
|||||
|||||
205963 AAGCATTCAGAGGTGTCTATAGAGAGAGTCTTGACTCTTGTTCAGAT 205914
607 PheGlyThrTrpGlySerAlaValSerAsnLeuPheTyrAlaHisAspSe 623
|||||
|||||
205913 CGAGGCTTCTGGCGCTGGGAGTCCCAAT...TTCTTAGATAAAGATAA 205867
623 rSerGlyLysProIleAspAsnTrpHisArgSerLeuGlyTyrLeuP 640
|||||
|||||
205866 GAAAGGGGAAAAA...CCCAATACCGTCATAAATCTGCTGCAATGTCTA 205820
640 heGlyIleSerThrHisSerLeuAspAspHisSerPheCysLeuAlaIa 656
|||||
|||||
205819 TCGGAGGTGACAGCGCAACTGTGTGAAACTTAATTAAGTTTGCCCTT 205770
657 GlnGlnLeuLeuGlyLysSerSerAspSerPheIleThrSerThrGluTh 673
|||||
|||||
205769 TGCCAACTCTTGTAGAGGATTAAGATTTCTTAGCGCTAAATAATCATAC 205720
673 rThrSerTyrIleAlaThrValGlnAlaGlnLeuAlaThrSerLeu... 688
|||||
|||||
205719 TGATACCTATGACAGAGGCTTCTATATCCAAACATTAACAGAAATGTAGTG 205670
688 ..... 688
205669 GGTTCATAGGTGTCTCTTAATAAATCTCTGGCTTGAGAGTCATAAA 205620
689 ...MetLysIleSerAlaGlnAlaCysTyrAsnGlnSerIleHisGluLe 704
|||||
|||||
205619 CCCCTCGTTTAGAAGGACAGCTCGCTATAGCCACGACATTAATGATCT 205570
704 uLysThrLysTyrArgSerPheSerLysGlyGlyPheGlySerTrpHis 721
|||||
|||||
205569 GAAGCAAGATATCTGCGTAT...CCYAGGTGAAGAAAGTTCTTGGGCGA 205523
721 eValAlaValSerGlyGluValCysAlaSerIleProIleValSerAsn 737
|||||
|||||
205522 ATATGCTTTTAACATGATGTGTGGAGCTTCTTCTCATTTTATCTTGAA 205473
738 GlySerGlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGlyPh 754
|||||
|||||
205472 TACGTGATGTTTGTGATACTATGCTCATACATCAAGATATGTCAGC 205423
754 eSerGlyThrGlnAspGlyPheGluGlnSerSerGlyGluIleArgSerP 771
|||||
|||||
205422 CTATATACGTGACAGACCTTCTCGGAAAGGTACAGAAAGAGATCTT 205373
771 heSerAlaSerSerPheArgAsnIleSerLeuProIleGlyIleThrPhe 787
|||||
|||||
205372 TTGATGACAGCAACCTCTTCAATTTATCTTGCCATRAGGGGTGAGATT 205323
788 GluLysLysSerGlnLysThrArgThrTyrTyrPheLeuGlnAlaTyr 804
```



```
/transl_table=11
/product="outer membrane protein 5"
/protein_id="CAB37071.1"
/db_xref="GI:4455886"
/db_xref="SPTREMBL:O956P2"
/transl_table="MKSPSWLVSSTLACFTSCSTVFATNTENITGSPDSFDCSTNTG
TYTPKNTTIGIDYTLTGDTTLQNLSDSAALTGCEPDSITDESIFAGKGSLSFLNKS
SAEGALSVTTDKNLSLTFSSSLTFLAASVITTPSGGAVKCGDLTFPDNNGTLF
KODYCEENGCAISTKNSLSTKNGSGISFPGNKSATGKGAICATGTVDITNNATP
LFSNNIAEANGCAISTGNCITLTGNTSLVFSNSNATNGGALISGDADYVITSGOS
VTFSGNAYANGCAIYAKKLTILASGGCGSISFNNIVQTTNGGALISILAGCESL
SVAEGDITFNGNAIVATTPQTTKRNISIDIGSTAKITNLAIISGHSIFEDYPTANTAA
DSTDTLNLKADANGSTVSGSIVFSGELSDKAVANLSTIKOPATLALAGNLV
KRGVTLDTKGTGTAGSSVIMDAGTTLKASTEEVLTGSLIVDSIGEBKKVYIAASA
ASKNVALSGPILLDNOGNAYENHDLGKODESPVOLSALGATTTDPAVPAVATPT
HYGOGTGWCTWVDPAASPRTKTKTALTAWNTGYLDPNPRGCRPIVPSNLSWSPVIOA
IOGVILERSALITLCSDRGFMAASVANTLDDKKGKERRKIKHKGCGAIGCAACOTCSNL
ISFAFQQLFSKDFLVAAKNHDTYAGAFYIOHTIECSGFICCLDLKLGSHKPLV
LEGOLAYSHVSDNLKTKYAYPEVKSGMGNNAFMNMLGSHSYEDELHCPDTYAPYI
KNTLYITRODSSEKGTGERSFDSNLFNLSPILGIYKFEKFSGDNDFSDLTLSVYPD
LIRNDPKCTTALVISCASMEVYANNIARQALQVRAAGSHVAFSPMEFVLQGFVEVRGS
SRIVYDLGCKRPF"
complement(11313..11372)
/gene="omp5"
11535..114321
/gene="omp4"
11535..11588
/gene="omp4"
11535..114321
/gene="omp4"
/codon_start=1
/transl_table=11
/product="outer membrane protein 4"
/protein_id="CAB37072.1"
/db_xref="GI:4455887"
/db_xref="SPTREMBL:O86164"
/transl_table="MKTSIPWVLSVSLAFSCHLOSLANELLSPDSEFNGNIDSGTF
TEKTSATYSLTGDFEYEPKGTPLSDSCFQOTDNLFLPFLNGSLTGTGFDAGHA
GAASTTANKNITFSGFSLSPDSPTVTGCGSLAGCVNLENIKRLVAVNFS
PADGAIKCAKSLTGTSGDALFSNNSSTKGCALATVAGIARVANNVCKLVANIAS
TSGCATDDECTSLISNNKFLYPEGNAKTTGCAICNPKSGPELISNNKLTIASN
VAETSGAIIHAKKLLSSGCTFEFLNNVSSATPKGAISTDSGELSISLETGNT
VANTLTGTSTDPKRNAINISNKEFELRAKKNITFEYDPIITSEGSVDLKTINN
GSAAGNLNPOGTLFSGEITLADDELKVNADLKSFTOPVLSGGKLLLOKGVLEST3
FSOAGSLGMDSGTTLSTTASGITITNGIWNDSGLKOPVSLAKGASNKVYVSK
LNLIDEGNIYESHMFSDOLPSLLKITVDADVDPNVDISSLIPVADPNSEYGEIG
QMNVTMTPTATNTKEATPTWTKTGVPSPERKSALVNCYILMGVTFDILSDQI.VEIG
ATGMEKOGCFWVSMTNFLHKTGDENRKGFRTSGVYIGSAHTPKDLFTFARCL
FARDKOCFTAINNNSRTYGGTLEFKSHHTLOPONYLRLGAKSESNAIEKFPREPLAL
DVOVSFSDNRMETHYTSLEPEGSMSNECIAGGIGIDLPLVLSNPHLFTFTFIOM
KVENYVSONSFESSSDGRGFSIGRLNLSIPVGAKFVQGDIGDSYTYDLSGFPVSD
VYRNPOSTATLVMSPDSWKIRGCLNSRQAFILRGSNNVYVNSNCELFCOHYAMELRGS
SRNVVDVGTCLRF"
14579..114644
/gene="omp13"
14579..16123
/gene="omp13"
14579..16123
/gene="omp13"
/codon_start=1
/transl_table=11
/protein_id="CAB37073.1"
/db_xref="GI:4455888"
/db_xref="SPTREMBL:O973D6"
/transl_table="MTILRNFLTCSALFLAFAAAOVYVLIHSDGYNCAIINKSLEPK
ITCYPEGTSYIFLIDVIRISNVKHDQADAVFINRSGNLSFEMKNRNFHTNMTBEGF
AIIISNRVGDYTLTNSFYLAFTSAPLFPOGQASVILFENRNSSEENVTFCGNYSW
SGAAIYTPYLLGSKASRPSVNLSGNRVLFVRNNGYSGAISTINLLITTGSGSFE
NNHAAHDVNSNGCAIAIARGCSISIVSKGDLIFGNGNASCNGNTIHNHILHOSGOF
KULRAVSEGVYFYDPISSHSESHKTTDLVYNAPECKTYEGTISFGCLDIDHEVCAE
NITSTILQDYTLACGTLISDVTYQLHSFRKBEASTITMSGCTLLCGCDRVONLH
IITEDTNEFVPIRAREDKDALVSLKLVAREAVASVYDFQEFKATPIPLELLGP
```

```
alignment_scores:
  Quality: 883.00      Length: 942
  Ratio: 1.541      Gaps: 31
Percent Similarity: 60.828      Percent Identity: 30.573
```

alignment_block:

us-09-677-752-4 x CPN133034/rev ..

Align seq 1/1 to reverse of: CPN133034 from: 1 to: 17280

```
12 CysAlaIaIaIeUeSerSerThAlaIaIeUePhcIyGlnAspProLe 28
|||||.....
11315 TGTTCACCTGTTTTCGTGCAACTGCAAAATATGAGCCCTCTGATAG 11266
28 u.....GlyClnThraIaIeUeThrIlyAsnProAsnIaIaIc 43
|||||.....
11265 CTTTGACGGAAGTACTAACACAGCCACTTACTCTCAAAAATACGACTA 11216
43 ys.....ThrPhehClnAspCysThrMetGluSerLeu... 54
|||||.....
11215 CTGCAATAGACTACTCTGACAGAGATATTACTCTGCAAAACCTTGGC 11166
55 ...PheProAlaIeUeCysAlaIaIaSerClnAsp.....AspProLe 68
|||||.....
11165 GATTGCGCAGCTTAAAGAGAGGTTTTCGACACTACGAAATCTT 11116
68 uTyValIeUeClnAspSerTyrcySTTPheValSerIyLeuHisIleT 85
|||||.....
11115 AACCTTGCCGTAAGGGTACTCA.....CTTCTTTTAAATATTA 11072
85 hrAspProLyGlnAlaIeUePhcIyGlnAspLeuSerIeGln 101
|||||.....
11071 ACCTACTGCGCAAGGC.....GCAGCACTTCTGTATCA 11037
102 AsnPhcArGpHeUeSerPheThAspCysSerIySglu.....Se 116
|||||.....
11036 ACTGATAAATCTGCTGCTAACAGATTTTCAGCTTACTTCTTAC 10987
116 rSerProSerIleIle.....HisGln 124
|||||.....
10986 GCGCCCATCATCGGTAATCACAACCCCTCAGAGAAAGTCCTTAAT 10937
124 yAsnGlnIyGlnLeuSerLeuArGAsnAsnGlySerMetSerPheCysArg 140
|||||.....
10936 GTGAGAGGAGATCTTACATTTGATTAACAATGCAACTATTTATTAACAA 10887
141 AsnHisAlaGluGlySerGlyAlaIaIeSerAlaAspAlaPheSerLe 157
|||||.....
10886 GATTACTGTGAGAAATGGCGGAGCATTTCTACCAAGATCTTCTTT 10837
157 uGlnHisAsnTyIeUePheThraIaPheGluGlnAsnSerSerIySgl 173
|||||.....
10836 GAAAAACAGCACAGGATCGATTTCTTTTGAAGGAATAAATCGACGCAA 10787
174 .....AsnGlyAlaIaIeGlnAlaGln...ThrPheSerLeuSer 186
|||||.....
10786 CAGGCAAAAAAGTGGCGCTATTCTCTACTACTGTAGATATTAACA 10737
187 ArgAsnValSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnG 203
|||||.....
10736 AATATATCGGCTCTACTCTCTCGAACAAATATTCGAAAGCTGACG 10687
203 yGlyAlaIle...CysCysSerAsnLeuIleCysSerClyAsnValAsn 219
|||||.....
10686 TGGACCTATAATACACACAGAAACTGTACAAATTAACGAAATACGCT 10638
219 rOLeUPhePheThrGlyAsnSerAlaThr.....AsnGly 231
|||||.....
10637 ..CTTGATTTTTCGAAATAGCTGTACACAGCAGCAGGAATGCGAGA 10590
232 AlaIleCysCylIeSerAspLeuAsnThSerGlyIySglSerLeuSe 248
|||||.....
10589 GCTCTT.....TCTGAGATGCTCGATGTATAC 10564
```


[illegible]

seq_name:	gb_bal:AE002192
seq_documentation_block:	
LOCUS	AE002192 12676 bp DNA BCT 30-MAY-2000
DEFINITION	Chlamydomonada pneumoniae AR39, section 25 of 94 of the complete genome.
ACCESSION	AE002192 AE002161
VERSION	AE002192.2 GI:8163405
KEYWORDS	

SOURCE	Chlamydomonadales; Chlamydomonadales AR39.
ORGANISM	Chlamydomonadales; Chlamydomonadales AR39.
REFERENCE	Bacteria; Chlamydomonadales; Chlamydomonadales; Chlamydomonadales.
AUTHORS	Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., 1 (pages 1 to 1267)

TITLE	JOURNAL
White, O., Hickey, E.K., Peterson, J., Umayah, L.A., Utefferback, T., Berry, K., Bass, S., Linner, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gilm, M., Nelson, W., Deboy, R., Kolonay, J., McClarty, G., Salzberg, S. L., Eisen, J. and Fraser, C. M. Genome sequences of <i>Chlamydia trachomatis</i> MOMP and <i>Chlamydia pneumoniae</i> AR39	Nucleic Acids Res. 28 (6) 1397-1406 (2000)

MEDLINE 20130235
 PUBMED 10684935
 REFERENCE 2 (bases 1 to 12676)
 AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
 White,O., Hickey,E.K., Peterson,J., Umagam,L.A., Utterback,T.,
 Berry,K., Bass,S., Linher,K., Weidman,J., Knouri,H., Craven,B.,
 Bowman,C., Dodson,R., Gwinn,M., Nelson,W., Debroy,R., Kolonay,J.,
 McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
 TITLE Direct Submission
 JOURNAL submitted (01-MAR-2000) The Institute for Genomic Research, 971
 COMMENT Medical Center Dr. Rockville, MD 20850 USA
 FEATURES On Jun 1, 2000 this sequence version replaced gi:7189226.
 source 1. .12676
 Location/Qualifiers

```

/organism="Chlamydomedilla pneumoniae AR39"
/strain="AR39"
/db_xref="taxon:115711"
/note="synonym: Chlamydia pneumoniae AR39"
complement(90..185)
/gene="CP0300"
complement(90..185)
/gene="CP0300"
/note="hypothetical protein; identified by glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"

```

```

/protein_id="AAF38157.1"
/db_xref="GI:7189227"
/translation="MFTSLGRSAIFYFTGSLKHLRLGEFLIFOK"
complement(275..1819)
/gene="CP0301"
complement(275..1819)

```

gene
CDS

```
/gene="CP0301"  
/note="similar to GP:4376735; identified by sequence  
similarity; putative"  
/codon_start=1  
/transl_table=11  
/product="polymorphic membrane protein G family"  
/protein_id="Aa038158.1"  
/db_xref="GI:7189228"  
/translation="MTILRNFLTSCALFLAIPAAQVYLHESDYGNGAINN  
ITCYEGEGSYTILDDVRISNMYKHDDAGVFINRSGNLEFPMNRCKETFEHNI  
AIIASNRQGTGTTLSNFSYLAFTSAPLIDPGGAGLISGVSIENSEETPCP  
SGAALTYVYLILGSKASRPSVNSGNRYVFNPSQGVGALISTNNLTITPTPL  
NGAIAHDVNSNGCALAIPAGSGTISIVYSGDILFRGNATASQDQNTIHSIHI  
KNLRVASEGVETFDPIISHSBKHITDLYINAPBEKREYEGTISFGCLDGLD  
NLSTISLDDVTLAGGTISLSDGVTQLQHSFKGEASSTILMSGTLILSGDAG  
ILILETDENVPIRAIEDKDALVSLSEKVAFEAAWSVYDFQFEALFIP  
FSDSLILETTLETRQVVTENDAVAGFWSLSEVEPPLSLDKDRITPTPKVY  
ETVSNP"  
complement(2077..4926)  
/gene="CP0302"  
complement(2077..4926)
```

```

/feature="CP0302"
/note="similar to GP:4376733, identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="polymorphic membrane protein G family"

```

/protein_id="AAF38159.1"
 /db_xref="GI:7189229"
 [translat] "M1YLFCEFYIDANSSLLKNSITMKTSIPVAVLSVLAFL
 ANEELLSPDSENGNIDSCGTFPPKTSATYTLGTGVGFEEFPGKGPVLSQSC
 LFLPFGNGSLTEFGFLDAGTHAGAASTAANKLLTFSGSLTSDSSPSSTY
 SSAGGVLENIEKFLVAGNCFSTADGGATLKGASELTLTGSGDLFENNSSTY
 TAGAGTANNQGVYFRLSNIASTGSGALIDEGSLTNNKFLFEENAKTTO
 KASGSEPLISNNKTLFTRSNVAFETSGCALHAKKALLKSLSGGTFEFLRNNSV
 ASIDASGSLSLSAETGNTFTRVNTLTGTTGDTFRKRAINIGSKGCFPEL
 LIFVYDITSEGTSLVSOVLKINNSAGALNLYOCTILPSCGTLADBLKVAIND
 PVSLSGGKLLLDQGVTLBSTSFSGAGSLLDGDSGTTLTGTGSLTGLNGL
 LKQPSLPAKSGASNVIVYSGNLNLDIEENIETSEIMFSDQLFSLKLTVDV
 DLSILIPVAPEDPNSKEYFGOGQNMVNTDTATNTKLEATATKGFSPV
 CMTLGVGTFDLSLQOLVELIGATGHEKHGQFVNSVMTNLHKTGDSRGGF
 VTGSGAHPKDDLPFPAFCHLPARDKDEIANNRSYVCGTLFPFKSHRPLQ
 GKPSSESIEKFEPRFELPALDLYVGFSDNSRMTHTYSLPEGSSNSPEFL
 LDLPVLSNPHRLPKFTIDQKMEVYVYQNSFESSSDGRCFSLGRLNLS
 FVQSGIDGYSYTTDLDSGFVSDYRNNPOSTATLVMSPSMKLTGGNLKQAA
 NVNYSNELFEHVIAMELGRSSRNINVDGTRLEP"
 5026..7812

CD5

/gene="CP0303"
5026..7812
/note="CP0303"
/note="similar to GP:4376729; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=1
/product="polymorphic membrane protein G family"
/protein_id="AA038160.1"
/db_xref="GI:7189230"
/translation="MKSQSWLVLSSTLACFPSCSTVEFAAENIGSDSFF
TTPNNTTGIDYLTGDTLTQNLGDSALTYGCPDSTGSSFPAGKGLSL
SAEGALSTYTDKLNLTGPFSLVFLAAFSVITTPSGKAVGCGGLDFD
KQDYENGGALSTKNLSLKNSTGSI SFGKNSMARTKGGALICATGPAIDI
FLSNNAIEAGAGINSTGCTGTGNTSLISFNNSVATARGNGALSGADVA
VTESENQAVNGAIYAKKTLTLAGSGDGSNNIVQGTGNGCAISIL
SAEADITFNKNAIYATTPOTTKRNSIDGSAKTNLNAIGSHIFRDP
DSTIDLTNFKRADNGSTDSGSIVSGEKLTSEDEKNAVNAIGSLIKQPTL
KGVGLTDPKGFVGTAGSSVITMDAGTTLAKSTAEVLTGLSIPVDSLGEGRK
ASKNAVLSPKILLLQNGAVNAEHLGKRDSPVQLSKATGATTTDDPAVAV

684LeuAlaThrSerLeuMetLys..... 690
 2174 TGCCCTCATTTGGAGATATACACCGACATGCGAAGATATTCCTCTCATTT 2223
 691 11SerAlaGlnAlaCysTyrAsnGlnSer11Leu15GlnLeuLysThrLy 707
 2224 TTGAATGCCAGCTAACTACACGCTACACTAAATVGTATGATGATCTCG 2273
 707 sTyrArgSerPheSerLysGlnGlyPheGlySerThrPheLysAlaVal 724
 2274 CTATACCTCTCAT...CCTGAAGCTCAAGGTTCCTGACCAATTAATTCCTG 2320
 724 a1SerGlyGluValCysAlaSer11LeuProLe...ValSerAsnGlySer 739
 2321 GGGCTCTAGAGCTCGAGCATCTCGCTCTATATCTCCCTAAACAAAGCA 2370
 740 GlyLeuPheSerSerPheSer11LeuPheSerLysLeuGlnGlyPheSerG1 756
 2371 CCCTCTCTCCAGGATATTCCTCCCTTTCCTTAAAGTTCACAGCAGCTACAG 2420
 756 yThrGlnAspGlyPheGlnGlnSerSerGlyGlnLeuLeuArgSerPheSer 773
 2421 CCGCCAAACAACTTAAAGAGAGTGGCCCTCAACGCCGCTCTTTCGATG 2470
 773 1aSerSerPheArgAsn11LeuSerLeuPro11Gly11LeuThrPheGlnLys 789
 2471 ATGAGAGACTAGTGAAGTCTATCCCTGTCGATTCGGTATGATAAAAA 2520
 790 LysSerGlnLysThrArgThrTyrTyrPheLeuGlnAlaTyr11G1 806
 2521 ATCTCCGAGAGTGAATAAATAATTCGAGATTCCTCTCAACCAACATTCG 2570
 806 nasPheLysArgAspVal1GlnSerGlyProValValLeuLeuLysAsn 823
 2571 TGATGCTGATGTAAAAATCCCGCTTCGCGATCTCTCAATGTCACGTC 2620
 823 1aValSerThrPaspAlaProMetAlaAsnLeuAspSerArgAlaTyrMet 839
 2621 GAGCTCTTGGACTTCGCTATGTAATAACCTCGCAGCAGACCTCTCTTA 2670
 840 PheArgLeuThrAsnGlnArgAlaLeuHis...ArgLeuGlnThrLeuLe 855
 2671 GCAACGCTGCAAGCCATCTGATCTCTCCCTCATCTCAACAACTCTCTCG 2720
 855 uasAlaSerCysValLeuArgGlyGlnSerHisSerLysSerLeuAspL 872
 2721 GGAACCTCTTATGAGCTTCGCTGCTGAGCACACTTACATGATGATTT 2770
 872 euGlyThrThrTyrArgPhe 878
 2771 GTGGCGTAAAGATACATCTTC 2790
 seq_name: gb_ba2:CP072499
 seq_documentation_block:
 LOCUS CP072499 4926 bp DNA BCT 13-MAY-1999
 DEFINITION Chlamydia abortus putative outer membrane protein gene, partial cds; and putative 98 kDa outer membrane protein gene, complete cds.
 ACCESSION U72499
 VERSION U72499.1 GI:1657776
 KEYWORDS
 SOURCE Chlamydia abortus.
 ORGANISM Chlamydia abortus.
 Bacteria: Chlamydiales; Chlamydiaceae: Chlamydia.
 REFERENCE 1 (bases 1 to 4926)
 AUTHORS Longbottom,D., Russell,M., Dunbar,S.M., Jones,G.E. and Herring,A.J.
 TITLE 98kDa protein genes from ovine abortion strain 526/3 Chlamydia psittaci
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 4926)
 AUTHORS Longbottom,D.
 TITLE Direct Submission

JOURNAL Submitted (25-SEP-1996) Moredun Research Institute, 408 Gilmerton Road, Edinburgh EH17 7JH, Scotland, UK

FEATURES

Source

/organism="Chlamydia abortus"

/isolate="ovine abortion 526/3"

/specific_host="Ovis aries"

/db_xref="taxon:83555"

/clone="S112"

/dev_stage="elementary body"

/note="submitted as Chlamydia psittaci"

<1.1950

/codon_start=1

/transl_table=11

/product="putative outer membrane protein"

/protein_id="AAB18187.1"

/db_xref="GI:1657777"

/translation="DPITTTAPAAAVVAKOPEASLAKATSGIPASGASVAPAP

FTPLKINADPDODPEIKVAAAOQSAVYNGKIYFSGEKLSSDPAKPLNATSVIH

NYSLFAGTLVSSGGLVDSFCTKEGSLIYMGDGTSTITVTPASELQSRTPPS

PKNAITVIRAVSKVYSSLLINRERADSGAVVPIIESPGSITITLAVNLDSLE

NGVITLAKGSGSVTLTGLOFQDOSQNFYDNPLNKNFTLNFLDLSAPAEK1HT

EEFNIIPOGATSSNLGOGKWEYEVKDSGVSFEKMYSTGYIPTARRATLVNS

WCSAIDMRAPFONLVEVSTEGDFHRGLMISG1SNFHKDSYKVOEGFPHISSGVV

VSTQPISSKRVMDLAFCOMLGSKSDYRLADARSHVYAASTHTCEKLVNHTPSKRG

ILARKPEKSPITFEDQLSYLSLNSMTTHTTPNPSRCGKWNHCYAGELGSLPTLYD

HAIEELFEFVLIHTFYNOEDPFKEKQGTENRNQSAIFVNSPLGRFETKLN

TNIRLAIOPDITRDAPKSKVPLPSVHTAMSTGATNLSKQAMILGSDHHLTDNLE

FCHGAFELGSSSNYNVDIGRKYF"

2083..2089

/note="putative"

/db_xref="GI:1657778"

/protein_id="AAB18188.1"

/product="putative 98 kDa outer membrane protein"

/translation="MRPSLYKILISSTLTLPISFPHSOLAARVALTOESIIIDNAGFS

POSTRAGCTIYNVESDIDVDVGOTAAALASSAPOTANLTFKGNHSLSTNNAG

ANPAGINVTADKILITLDFSKLSRECEPSSLVNTGKMGKSGALNLNANNS1LEDO

NSAENGAIISCKAPSLTGSKEISFTTNSIAKGAIAIGAILSDNOGTIRPSGN

TVNSGCAVYSEASMTIACNNHVAFAFNNAVSCSDCGAIHQSGTGAPLTIRDNK

VLFEENTSSAKGAIYTDKILITSGCPAPFINNKTHATPAGAIAGIANEGCSLTA

EHGDIITFONNLATODNATIKRNAIINIEBNGFVNLRASGKTSIPYDITVEGNAD

LTINKAREGDKTYNGRITISGEKLEPEQAAVADNLTTPDTQITTLAAGLVRSGEV

EAKTVYVOTGSLIMDAGTKLSAKTEDATLTLALPNTLDGKRAVDAAVAGNVT

LSGALGVIDPTDPEYENHLLNDTLALGGLISGKGSVTTNVPSHVGALEHYGOG

MSVSVMKDNNSDPKTOA1FTWNTKTVYVNPBRAPVLVNLMSFIDLRSLIOVLE

RSVDSLLETRGILVSGICGNPFHKDINAERKFRHISGVLGATNTSRSDLSVAF

COLFAKDKDYLVSKNAANYAGSVYVOHNSKRFDDLTLRLFGNPNTCCSGSKRPLIFLD

NOITYCHTANNMNTSTYTDPEYKSGNNTLCLITLSTYPIFYVSSSITDSTAPAKL

QVYVYAHQDDPKEPTTEGRVSSDLNLSVPIGIRFERKLSVGERSAVDLTLLVYIDVY

RNPSCMTGLAINDVSWLTATNLAARQAFIVRAGNHIALTSGVEMFSQGFELRSSSR

NYNVDLGAARVAF"

4884..4905

/note="potential 9-base terminator sequence"

stem_loop

BASF COUNT 1472 a 1033 c 1031 g 1390 t

ORIGIN

alignment_scores:

Quality: 896.50 Length: 848

Ratio: 1.731 Gaps: 25

Percent Similarity: 61.085 Percent Identity: 30.307

alignment_block:

US-09-677-752-4 x CP072499

Align seg 1/1 to: CP072499 from: 1 to: 4926

98 LeuSer11GlnAsnPheArgPheLeuSerPheThrAspCysSerSerLy 114
 |||||
 2442 CTACAGCTGACAGATTTTCTAAGTTCGATTAAGCAATGCCCATCTC.. 2489

```

13660 CCCCTAGACGGAATTAAGAGTGTCTCCATTCGAACACGAGGACCCACAA 13709
486 .....AsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAsp 500
13710 AACATTAACCTTAACCTCTCTCTCTTTTCCAGATAGTAGCGCG... 13755
500 LuasnPheTyrgluAsnValGlnLeuLeuSerLysGlnGlnAsnAsnIle 516
13756 ..AATTTTATGAAGAGCATAGATAAACCAAGCCTTCACCGACCTTTG 13803
517 ProLeuLeuThrLeuSerLysGlnGlnSerHisLeuHisLeuProAspG 533
13804 GTGGATATTCACTGCTGCTACACGCTACGCAATTCATTTATATC... 13850
533 YasnLeuSerSer.....HisPheGlyTyrglnG 543
13851 GCCTTCCTACTCTCCAGTACAAACCTCAGAACCTCATACCGGTACAG 13900
543 YAspTrpThrPheSerTrpLysAspSerAspGlnGlnHisSerLeuIle 559
13901 GACATTGGGAAGCCACTTGGGACAGACATCACTGCAAAATTCAGAACT 13950
560 AlaAsnTrpThrProLysAsnTyValProHisProGlnuArgGlnSerTh 576
13951 ATGACTTGGGTAACCTAGCGGCTACACCCCTAATCCTCAGCGTACACCT 14000
576 rLeuValAlaAsnThrLeuTrpAsnThrTySerAspMetGlnAlaValG 593
14001 CGTGGTTCGCCATTCATATGCGCATCTTACTGACATTCGCACTGTAC 14050
593 InsertMetIleAsnThrIleAlaHisGlyAlaIleTyLeuPheGlyThr 609
14051 AGCGATCATCATCATCAAGCAAGATAGTATCTTCAGCAACAGAGGACTC 14100
610 TrpGlySerAlaValSerAsnLeuPheTyValHisAspSerSerGly 626
14101 TCGGCATTCAGCACTGCCAATTCCTTCAT...AAGCATAAATCAGGAC 14147
626 sProIleAspAsnTrpHisIleArgSerLeuGlyTyLeuPheGlyIle 643
14148 TTAAC...CAACGATTCGACATAAAGCTACGCTATATTGTGGAGAA 14194
643 erThrHisSerLeuAspAspHisSerPheCysLeuAlaIleGlyLeu 659
14195 GTGCTGAAGATTTTCTGAANAATCTTCAGCTAGCTTTTGGCAGCTC 14244
660 LeuGlyLysSerSerAspSerPheIleThrSerThrGlnThrTrpSerTy 676
14245 TTCGGTAAAGATAAAGACCTTTATAGTTGAANAATACCTCTCATACTA 14294
676 rIleAlaThrValGlnAlaGln..... 683
14295 TTTAAGCGTCTAATACCTGCAACATCGAGCATTCCTAGGAGACTTCCA 14344
684 .....LeuAlaThrSerLeuMetLys..... 690
14345 TGCCCTCTTTGGAGATATCAACGCAATGCTGAAGAATATTCTCTCAT 14394
691 lIleSerAlaGlnAlaCysTyTrpAsnGlnSerIleHisGlnLeuLysTrp 707
14395 TTGCATATGCCAGCTAAGCTACACCTAAATCATATATGATATACGCC 14444
707 sTyTrpArgSerPheSerLysGlnGlyPheGlySerTrpHisSerValAla 724
14445 CTATACCTTCCTAT...CCTGAAGCTCAAGGCTTCGACCAATTAACCT 14491
724 alSerGlyGlnValCysAlaSerIleProIle...ValIleSerAsnGly 739
14492 GGCCTTACAGCTGGAGGATCTCTGGCTCATATCTCTCTTAAGAGACA 14541
740 GlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGlyPheSerG 756

```

```

14542 CCGTTCTCCAGGATATTTCCCTCTCTTAAGTTCAGGAGCTACAG 14591
756 yThrGlnAspGlyPheGlnGlnSerSerGlyGlnIleArgSerPheSer 773
14592 CCGCAACAAACATTTTAAGAGAGTGGCGCTGAAGCCGCTTTTGATG 14641
773 laseSerPheArgAsnIleSerLeuProIleGlyIleThrPheGlyLys 789
14642 ATGAGACCTAGTGAACCTGCTATACCTGTGCGGATTCGGTTGAAA 14691
790 LysSerGlnLysThrArgThrTyTrpTyTrpPheLeuGlyAlaTyrlle 806
14692 ATCTCCAGATGAAAAAATAATTCAGATTCCTGACCTACATTCG 14741
806 nAspLeuLysArgAspValGlnSerGlyProValValLeuLeuLysAs 823
14742 TGATGTATATCGTAAATGCCCCGCTGCTGCTATCTGATGTCAGT 14791
823 lavalSerTrpAspAlaProMetAlaAsnLeuAspSerArgAlaTyMet 839
14792 GAGCCTTTGGACTTCGCTATGTAAACCTCGCAGACAGAGCTTTTA 14841
840 PheArgLeuThrAsnGlnArgAlaLeuHis...ArgLeuGlnThrLeu 855
14842 GCAAGTGTGGAACCCATGAGCTGTCTCCCTCATGTAGAACCTGTG 14891
855 uAsnValSerCysValLeuArgGlyGlnSerHisSerTySerLeuAsp 872
14892 GGAAGCTGCTTATGAGCTTCGTGCTGACGACACATCTCAATGTA 14941
872 euGlyThrThrTyArgPhe 878
14942 GTGGGCTAAGATCTCATTC 14961

seq_name: gb-pat1:A81841
seq_documentation block: 2793 bp DNA PAT 21-JAN-2000
LOCUS A81841
DEFINITION Sequence 15 from Patent WO9858953.
ACCESSION A81841
VERSION A81841.1 GI:6731871
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 2793)
AUTHORS Madsen,A. and Birkelund,S.
TITLE NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE
JOURNAL Patent: WO 9858953-A 15 30-DEC-1998;
MADSEN ANNA SØFTE (DK); BIRKELUND SVEND (DK)
FEATURES
source
1..2793
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 803 a 681 c 567 g 742 t
ORIGIN
alignment_scores:
Quality: 906.50 Length: 840
Ratio: 1.691 Gaps: 26
Percent Similarity: 63.810 Percent Identity: 30.714
alignment_block:
US-09-677-752-4 x A81841
Align seg 1/1 to: A81841 from: 1 to: 2793
98 LeuSerIleGlnAsnPheArgPheLeuSerPhe.....Th 109
355 CTAACTTCAACAGATTTTCTAACCTTCTCTCATTCACACTCTCGAAC 404
109 rAspCysSerSerLysGlnSerProSerIleIleHisGlnLysAsnG 126

```

SOURCE Chlamydomophila pneumoniae CWL029.
ORGANISM Chlamydomophila pneumoniae CWL029
REFERENCE 1 (bases 1 to 15068)
Bacteria: Chlamydiales; Chlamydiaceae; Chlamydomophila.
AUTHORS Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W.,
Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)
MEDLINE 99206606
PUBMED 10192388
REFERENCE 2 (bases 1 to 15068)
Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,
Grimwood, J., Davis, R.W. and Stephens, R.S.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University
of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
FEATURES
source
1. 15068
/organism="Chlamydomophila pneumoniae CWL029"
/strain="CWL029"
/db_xref="taxon:115713"
gene
99.. 626
/gene="Cpn0439"
99.. 626
/gene="Cpn0439"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD18583.1"
/db_xref="GI:4376722"
/translation="MSTLNGVFPSPSLPEESADLFITNKEIVLAGEKNVFTLHSIPM
HIAITLIVLVALGAIICLCYSOSILLIIVGIVLTLLCLQALVGFKEFIROL
IIOOLHTTVQPIREKIRPESSIOLVINAKKTTQDILKLYEELCDLISQKFKIQSITLYO
KKELSHKKEKTNON"
651.. 1289
/gene="Cpn0440"
651.. 1289
/gene="Cpn0440"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD18584.1"
/db_xref="GI:4376723"
/translation="MATSVAPEPPESSPLSHATEVLNPNAYITOPHPIPAAMETF
SEKLSKRLHLCALITLLIGCTISAGVAGYGNMIIIGIGIIVLTLLALLAIP
LKKNOTGKTLIDEISODISSISCPQHYGLMPSHTKSKVHPDELFTIONOEKRLINEI
EKKESIOHLELKITTECONKLAKQOPKRSKSKSTKSKIKLSKNFVILPDC"
1437.. 2387
/gene="Cpn0441"
1437.. 2387
/gene="Cpn0441"
/codon_start=1
/transl_table=11
/product="CT007 hypothetical protein"
/protein_id="AAD18585.1"
/db_xref="GI:4376724"
/translation="MFKLFHIAFAFGHVLSPPIFYVADAGCIDEACKNPPRPESA
OYOYLKVNDAKFKLPHOITGYROYDGFCTCLPITPHKSSILLFSTCYICADIDKMSSSL
PISTDPNGLCNATFODTSFYNYVLISLIGATYLSLKNQMSIILSLVDPKNIENKYS
LVQVLSGKYGATEKLSAIFGVINEGCLHOKRAMPVGVSKATDQTLNLCYIPVVS
IDYRSTVCNCLAYRLTFRRKRLKLNHLLISSRGIFEYOGREIEANVKLTLPWGSFIK
GFYSGICNDISLADHNNKTSHTFKTSAFGGSAYVMNF"
2559.. 3077
/gene="Cpn0442"
2559.. 3077
/gene="Cpn0442"
/codon_start=1
/transl_table=11
/product="CT006 hypothetical protein"
/protein_id="AAD18586.1"
/db_xref="GI:4376725"
/translation="MGFKNICOGSOLYNGIPPERIILARKLKNCKSPRIALTIEV
LVSSVIGALKVILIPCASTYVALTLPLRALFNAIKTKSCOHILASYAMAMLHLITIAV

gene
3275.. 4528
/gene="Cpn0443"
3275.. 4528
/gene="Cpn0443"
/codon_start=1
/transl_table=11
/product="CT005 hypothetical protein"
/protein_id="AAD18587.1"
/db_xref="GI:4376726"
/translation="MSOPINPLGQPOVPAAPSPGQSPVYKRLKTSSTGLKFRPIT
PDYKPKRMRYVDGTGIIALAIILSLTLTASGNSLMLYALAPALAGVALLISDI
LDSPKARKIGEITAIWPIIYLAIAAGIACAFVASSGTMVLPMPFVGLITVGL
YFMSLNKLTIDVFRREHRLRMEKTOPTAEPIILVPSADMAKILAVEKQDLASAM
EHEKASODAHARRICREACOSPFYSSRNPHRKSFGSLSEFKKPSDASTASAI
SPFRKDDQPIHFKDLRSSFGSGASAFPIPMPSSPSPNSTGTIVLHPEVYPRKG
KEPSIPRVSSSSRSRPRDQKQ00Q00Q00DEEQK00K00KSKSKS00SLTTPPDKST
ANLSPSPFSSDEREKRRKRNK"
4833.. 9056
/gene="pmp_6"
/note="Cpn0444"
4833.. 9056
/gene="pmp_6"
/codon_start=1
/transl_table=11
/product="Polymorphic Outer Membrane Protein G/I Family"
/protein_id="AAD18588.1"
/db_xref="GI:4376727"
/translation="MKYSLPWLITSSALVFSLHPILMAANTDLSSDNYENGSSAFA
TAKESVDSAGTITLSDVSIINVAIIPADKSCFTNNGALSEFGADHSILQIHAL
THDGAIINNTNALSFEFGSSLLIDSAFATGSGGKAIIVTNEGATFEDNMSVT
LOKNTSEKMGALVSAVSIIDAKTTTAAIADONTSTKNGGALCSTANTTVQSGSVTF
SNTWATDKGCGIYSKEKSDTIDANTGVVTFKNTLPFGNMTAEFGAICATV
ENKTTGSAOANNPSCGCAICCYLATATDKGLAISQNEKSFISNTTANGCAIYA
TKCTLDGNTLTFTDONTATAGCGAIYETEDFSLKSGTIVTFSTNKTATGALYSK
GNSSLGTNNTLSEGNKATGPNSSNANOGCGGALAFIDSCVADKGLSTANNOEV
SLTSNAATVSGGAIYATKCTLTGNSGLFEDGATAGSGAIYETEDFTLGTSTV
FSTNATKAGALYSKGNSSLSGNTMLTFSGNKAQPSNSANOGCGGALISFLSAS
VSTKKGKIMEDNENVSLSGNTATVSGGAIYATKCAIHCWNTLTLPFGNMTAEFGA
ETEDFTLGTSTVTFSTNKTATGALTKGNTSTKKAIFYSNSKATATATTTTDO
ECCGGAILLNISDIDATKSLTLENESLFTINNTAKRSGGIIIVAKCVISSESTNF
DONTAETSGGAIYSKRLSTANGPVSFTNNSGCKGAIYASGSLSEALIDGDTFS
GNRATEGSTPNSIHLGAGATITKLAAPGHTIYEDPTMKPAGSGTIEELVYINP
VKAIVPPQPKPGPIASVVPVAPANPAGTIVVSGGLTIPDQASIPANTTTINPK
IILACGNVAKKSGATLOVSPFOGDPDSVTFQAGTILETTTNNNDGSDIKNLISVL
DALDKRMATTVIYVNSTSGSLKISGDLKPHNNKSGSYDNGKANNLNPFLDLSSTST
VALDQNPILPSSMAAPDITGIGSWTLVPEVGAAGVTLVAEQALGITPKPELRATLV
PNSLNNAVNIHISIOEITATMSDAPSHGIIIGIIGNAFHODKQKENAGFRLISRGY
IVGSGMTPEOETTFVAFSOLFGSKSDYVSDIKQVYAGSLCAOSSVYIPLHSSLR
HVLKSVLPPELPGETPLVHGOVSYGRNHNMTKLANNTOGKSDMSHSFAYEVGSL
PVDLNVRYLTFSYSPVYKLOVSNKQGOEVAADBRIDASHLVNVSIPMGLTFKHES
AKPPSALLTTIGYAVDAYRDHPHCLSLTNGTSNKTFRATNLSROAFPAASGLKLLH
GLDCAFGSGCELRSSSRSYNANGTRYSE"
9318.. 12128
/gene="pmp_7"
/note="Cpn0445"
9318.. 12128
/gene="pmp_7"
/codon_start=1
/transl_table=11
/product="Polymorphic Outer Membrane Protein G Family"
/protein_id="AAD18589.1"
/db_xref="GI:4376728"
/translation="MKSSYSMLFSSSIPLFSSLSIYAARVTLDDSSNNSVDSNGTFT
VFSTIDAAAGTIVSLSDVSTFONNAGHLIPLASGCPLEAGDLTTOGNOHAKRFIN
AGSSAGTAVASTAADKNLLEFNDFSRLSITISCSLLSPGOCALSVGLSLTNGSNI
IPTONFSSDNGCVINTKNTFLSGTQFAFSFNGAFGQGVAVATGTTIENSPGI
VFSFONLAKSGCALYSTDNCSITDNFQVYIPGNSAMEAQAQACATCTTDTKIVTL
TKNKLISFTNNALTYGCAISGLKVISISAGGPTLLOSNTSGSSAOGGAGGATVINSAG
ELASATSQSDTIFNNNOVNGCSTSRNAINIIDTKAVSIRARATSGYFVDPITNPG
TAASDTLMLNLADANSEIEYGAIFYSEKSLSPTEKALAAVNTSTIRQPAYLAAGDL
VRDQVTVFKDLTOSPGSRILMDGOTTLTSAKEANLSLNGLAIVNLSLDGINKAALKT

gene
CDS
complement(2941..5751)
/gene="CP0308"
complement(2941..5751)
/note="similar to GP:4376728; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="polymorphic membrane protein G family"
/protein_id="AAF38165.1"
/db_xref="GI:7189236"

ILNAQLSYSTKNDMDTRYTSYPEAOGSMWNNSGALLEGSLATLYLPKEADPEQGYFP
FLKFOAVYSROQNEKSGEAREADDGLVNCISIPVGRLEISDEKNEEISLAYI
GDVYKRNPRSRISLWVSGASWTLCKNLRAQAFILASAGSHLTLSPHVELSGAAVELR
GSAHLYNDCGLRYSF"
complement(2941..5751)
/gene="CP0308"
complement(2941..5751)
/note="similar to GP:4376728; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="polymorphic membrane protein G family"
/protein_id="AAF38165.1"
/db_xref="GI:7189236"

BASE COUNT 2726 a 2051 c 2367 g 2882 t
ORIGIN

alignment_scores:
Quality: 919.50 Length: 840
Ratio: 1.712 Gaps: 26

Percent Similarity: 63.929 Percent Identity: 30.952
alignment_block:
us-09-677-752-4 x AE002193/rev ...
Align seg 1/1 to reverse of: AE002193 from: 1 to: 10026

98 Leuser1leGlnAsnPhaArgPheLeuSerPhe.....Th 109
|||:||||: |||:|||||
2543 CTAAACATTCACAGCATTTTCTACCTTCCTTCATTCGACGCTCGGAC 2494
109 rAspYSerSerLysGluSerSerProSerIleIleHISGlnLysAsnG 126
||||: |||:||||: |||
2493 TACAGTGTCTTCAGAAAAGTACTTAAGT.....TCGCGAG 2456
126 LysGlnLeuSerLeuArgAsnGlySerMetSerPheCysArg..... 140
|||:||||: |||:||||: |||:||||: |||
2455 GACCTTAATCTTACCGATATGGAACGATCTCTTACGCCAAACGTC 2406
141 ...AsnHISlaGluLysSerGlyGlyAlaIleSerAlaAspAlaPhe 156
|||:||||: |||:||||: |||:||||: |||:||||: |||
2405 TCCAAATCAACCTAATAACATGCGGACGATCAGCCGCAAAACCTTTC 2356
156 rLeuGlnHISAsnTyLLeuPheThrAlaPheGluGluAsnSerSerLysG 173
||||: |||:||||: |||:||||: |||:||||: |||
2355 TATTCTGGGAATACCTCTTCTAATACCTTCACTAGTAATAGCGCAAAA 2306
173 LysAsnGlyGlyAlaIle...GlnAlaGlnPheSerLeuSerArgAsn 188
||||: |||:||||: |||:||||: |||:||||: |||
2305 AATTAGTGGAGGATCTATAGCTCTGCGGCTCCAGATATTAGGAAC 2256
189 ValSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnGlyLys 205
|||:||||: |||:||||: |||:||||: |||:||||: |||
2255 ACCGCGCAGTTAGCTTATGATTAATAAGAGAAAGTGGGTGGGC 2206
205 aIleCysCysSerAsnLeuIleCysSerGlyAsnVal.....Asp 219
||||: |||:||||: |||:||||: |||:||||: |||
2205 TCTG.....GGCTTGAAGCCACCTCTCGATTACTCAAAATAGCT 2165
219 rLeuPhePheThrGlyAsnSerAlaThrAsn.....GlyGly 231
||||: |||:||||: |||:||||: |||:||||: |||
2164 CCCTTTCTCTCTCGAACAACCTGCAACAGATGCTGAGCAAGGGCGG 2115
232 AlaIleCysCysIleSerAspLeuAsnThrSerGluLysGlySerLeu 248
||||: |||:||||: |||:||||: |||:||||: |||
2114 GCCATTATATGCT.....GAAAAACAGAGAGAGCTCTACTCTTAC 2074
248 rLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysGlu 265
||||: |||:||||: |||:||||: |||:||||: |||
2073 TACTCTGGAATAAAGTGTGACCTCGCGGAGAACTCTTCACTAATC 2024
265 YsgGlyAlaIleTyAlaLysHISMetValLeuArgTyAsnGlyPro 281
||||: |||:||||: |||:||||: |||:||||: |||
2023 AAGCGGAGCAATCTGTGCCATGCTGTAGATCTTCCGCTGCTGGCCT 1974
282 ValSerPheIleAsnAsn.....SerAlaLysIleGly 293
|||: |||: |||: |||: |||: |||: |||: |||
1973 ACCGATATTTCATATATATAGTCGCGGACACAGCTGCACGCAAGGGCG 1924
293 yAlaIleAlaIleGlnSerGlyLysSerLeuSerIleLeuAlaGlyGlu 310
||||: |||: |||: |||: |||: |||: |||: |||
1923 CGCATTCGATTCGCGACCTCTGATCTTAAAGTCTCTCTGCAAAATCAG 1874
310 LysValIleuPheGlnAsnSer...GlnArgThrSerAspGlnGly 325
||||: |||: |||: |||: |||: |||: |||: |||
1873 GAGCATTCACGTTCTTGGCAATACCTTAACCTTCACCTCGGCGCAACA 1824
326 LeuValArgAsnAlaIleTyLLeuGluLysAspAlaIleLeuSerSer 342
|||: |||: |||: |||: |||: |||: |||: |||
1823 TCGACACGAGATGCTATCTACTCTGATCGCAACAAATTAAGAACTT 1774
342 uGluAlaArgAsnGly...AspIleLeuPheAspProIleValGlnG 358
||||: |||: |||: |||: |||: |||: |||: |||

Quality: 940.50 Length: 910
Ratio: 1.761 Gaps: 32
Percent Similarity: 58.681 Percent Identity: 31.538

alignment block:
US-09-677-752-4 x AB1839 ..

Align seg 1/1 to: AB1839 from: 1 to: 2787

```
85 ThrSprProLysGluAlaLeuPheLysGluLysGluLysLeuSerIleG1 101
|||||.....
154 ACNAATGACATGACATCTATATATCTAAGAGGGATGCTCAATAC 203
|||||.....
101 nAsnPheArPhe...LeuSerPheThrAspCysSerSerLysGluSerS 117
|||||.....
204 CAATGCAAGATCTCCGACACGCTTAACCCCTCTGCTTAAGAACTA 253
|||||.....
117 erProSerIleIleHisGluLysAsnGly...GluLeuSerLeuArGAsn 132
|||||.....
254 CTGGGAATCTTTCTTCCAAAGCCAGCTACCAATTTCTCTCAAAAT 303
|||||.....
133 .....AsnGlySer 135
|||||.....
304 ATCGATCGGGAGCGAAGCTGACTTACCAATACAGCTGCAAAATAGCT 353
|||||.....
135 rMetSerPhe.....CysArGAsnH 142
|||||.....
354 TCTCTCCTTTGACGATTCCTCTATTTGTCTATATACAAACGAGATG 403
|||||.....
142 IsAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSerLeuGln 158
|||||.....
404 CTACCACAGGAGACGAGCCATCAAGTCCACGAGAGCTTTCTATTCG 453
|||||.....
159 HisAsnTyLeuPheThrAlaPheGluLysAsnSerLysGluYAsnG1 175
|||||.....
454 TCGAATCT...AGTGTCTACTTTGGCCAAACCTTTCTATATGCAATGG 500
|||||.....
175 yGlyAlaIleGlnAlaGlnThrPheSerLeuSerArGAsnValSerProI 192
|||||.....
501 AGCGCCCTCCCAAGCAAGCTCTATACGTATCTTAACCCCA...C 547
|||||.....
192 LeSerPheAlaArGAsnArGAlaAspLeuAsnGlyGlyAlaIleCysCys 208
|||||.....
548 TAACGTTTGGCAAAACAAACGCAACGCAAAAGGGGGCTCTATATCC 597
|||||.....
209 SerAsnLeuIleCysSerGlyAsn...ValAsnProLeuPhePheThrG1 224
|||||.....
598 ACGGAGGAGATTACATTAACATAACGTTTAACCTCAGCATCTTTCTGA 647
|||||.....
224 yAsnSerAlaThrAsn..... 229
|||||.....
648 AAATACCCGCGCAACATGGCGGAGCCATTACAGGAGACTACCACTT 697
|||||.....
229 ..... 229
|||||.....
698 TTATTAGCAGCAACAAAGCAATTAGCTTTATTAACAATAGTGTGACCGCA 747
|||||.....
230 .....GlyGlyAlaIleCysCysIleSerAspLeuAsnThrSe 242
|||||.....
748 ACCCTGACGACAGAGGGAGCCATTACTGTAGTAGT.....ACATCGAGC 791
|||||.....
242 rGluLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAlaS 259
|||||.....
792 CCCCAAACCACTTAACTTATCAGACACGCGGAAGCTGAACCTTATAG 841
|||||.....
259 erAsnSerAlaLysGluLysGlyAlaIleTyAlaAlaLysHisMeVal 275
|||||.....
842 GAAATACAGCAATTAAGTGTGGGGGATTTATACGACAATCTACTGTT 891
|||||.....
276 LeuArGTyAsnGlyProValSerPheIleAsnAsnSer..... 288
|||||.....
892 CTTTCTTGTGAGAGACTACGCTTTTAAACAACTCTGCTATAGATAC 941
```

```
289 ...AlaLysIleGlyGlyAlaIleAlaIleGlnSerGlyLysLeuS 304
|||||.....
942 TGCAGCTCCCTTAGAGAGAGCAATTCGATTCGACTCGATCTTGA 991
|||||.....
304 erIleLeuAlaGlyGluGlySerValLeuPheGlnAsnAsnSer..... 318
|||||.....
992 GTCTTGGCTCTTGTGGAGACATCACTTTGAAGGAAACACAGTAGTC 1041
|||||.....
319 ...GlnArGThrSerAspGlnGlyLeuValArGAsnAlaIleTyLeu... 333
|||||.....
1042 AAAGAGACTTCTTGACCTCAGACACACTACCGAAGAAATCTATTAACTCCG 1091
|||||.....
334 .GluLysAspAlaIleLeuSerSerLeuGluAlaArGAsnGlyAsp...I 349
|||||.....
1092 AAACACCAATGCTTAAGATTGATGACGTGCGAGCCTCAAGCAATTA 1141
|||||.....
349 LeLeuPhePheAspProIleValGlnLysSerSerLysGluSerPro 365
|||||.....
1142 TCTACTCTATGATCTTATTAACACTACACT..... 1176
|||||.....
366 LeuProSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAl 382
|||||.....
1177 .....CCAGCTCTCAGATCTCTAAACTTAATGC 1208
|||||.....
382 aSerProLeuValIleGlnThrSerAlaAsnArGSerValIlePheSerS 399
|||||.....
1209 TCTTGACCTTCGACGAGAACTCCGATATCAAGAACATCGATTTCTG 1258
|||||.....
399 erGluArGLeuSerGluGluLysThr...ProAspAsnLeuThrSer 414
|||||.....
1259 GAGAGAGCTTCGGAAGCAGAACCTCAGAACCTGATTAATCTCAATCT 1308
|||||.....
415 GlnLeuGlnGlnProIleGluLeuLysSerGlyArGLeuValLeuLysAs 431
|||||.....
1309 ACAATTACGACCACTTAACCTTTCGCGGAGGCACTCTCTTAATC 1358
|||||.....
431 pArGAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeu 448
|||||.....
1359 AGCAGTCACTCTAGCTCTTAAGCTCTTTCCCAATCCCGCGCTACCC 1408
|||||.....
448 euIleMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeu 464
|||||.....
1409 TCCATGATGAGGAGGAGCCACATTAGAACCCTGATGGATCTACTATC 1458
|||||.....
465 AlaThrLeuSerIleProLeuHisSerLeuAspThrGluLysSerValTh 481
|||||.....
1459 AATATATCTTTGTCATGTAGATCTCTTAAGAAGACCAAGAGCCTAC 1508
|||||.....
481 rIleHisAlaProAsnLeuSerIleGlnLysIlePheLeuSerAsnSer. 497
|||||.....
1509 GCTAAAGACCAACACAGCAAGCT...CAGACAGTCACTTATCTGATCGC 1555
|||||.....
498 .....GlyAspLysAsnPheTyGluAsnValGluLeuLeu 509
|||||.....
1556 TCTCTCTGTAGATCTCTGAAATGTCTAGCAAGATGT..... 1596
|||||.....
510 SerLysGluGlnAsnAsn.....IleProLeuLeuThrLeuSerLy 523
|||||.....
1597 .....TCTTGGAATTAACCTCAAGCTTTTCTGTCTCACTCTTACTGC 1640
|||||.....
523 sGluGln...SerHisLeuHisLeuProAspGlyAsnLeuSerSer... 537
|||||.....
1641 TGACGACCCCGCAATATTCACATCAGAC.....TTACTCTGTATC 1684
|||||.....
538 .....HisPheGlyTyGlnGlyAspTrpThrPhe 547
|||||.....
1685 CCCTAGAAAAAATCTATCTATGGGGATGCAAGGGAATTTGGGCATTA 1734
|||||.....
548 SerTrpLysAspSerAspGluGlyHisSerLeuIleAla.....AsnTr 562
|||||.....
1735 TCTTGCCAGAGATCTGCGACATAATCCAAACGACGACTCTTACTCG 1784
|||||.....
562 pThrProLysAsnTyValProHisProGluArGlnSerThrLeuVala 579
```

```
gene complement(8391..10808)
/ gene="gyrB_1"
/ complement(8391..10808)
/ gene="gyrB_1"
/ codon_start=1
/ transl_table=11
/ product="DNA gyrase, subunit B"
/ protein_id="BA98485.1"
/ db_xref="GI:8978649"
/ translation="MDPKENYDAATVLEGLQAVRPPGMVIGDTGITLHLIYVE
LVNDSIDAMAGKSRIDRILEDGGIVTVNGRGIPIEVEHRESAKGRVSALEYV
LTVLHACKKPKDKSKYKSCGLHGVCSVNAISELVAVVKKDKKQOMERSRCIPV
PLOTVSVSDROGTEIVFPDPKISTCTCFDPSILMKRLBELAFILNRGITIYFEDRDV
SFDEVTFEYEGGIQSFVSYLNQKESLFSFPIYICGTAVGDDGI EFEALQWNSCY
ELVSYANNIPTRGQTHLTGFSTALTRINTYIICAHNLAKNNKLTALGEDIREGLTA
VISVKNPPOEGOTKOKLGNSDVSVQAQVGEALITFEENPQIARMIYDKVAA
OAREAKKARELTLRKSLDLSARLPGLIDCLEKDPRECEMYIVEGDSAGSAGKGRV
RRFOALIPRIGKILNVEKARIKIQNOEIGITIAALCCGICADQNFNLSKIRYRILTI
MTDADVDSHIRTLLITFTYRHMALITNECVYIAQPLKYKSKKDRYILSEKMD
SYLLMLGTNESSLEKSTERELRGALLESFINLVLDVESFTNLEKKAIPSEFLEMV
KEGIGYLYVLAIPATMGOGRYLYSDEKEEALQDETHKRIIELYVVAEVDIQNG
LKEYGLDISYLLIPQKNEIVIGNEDSPSCNCTYLEVINLKLGRKGIETIORYK
LGENNADQLMDPTMNEQRTLIHVS LKDAVEADHIFTMLMEVEVPREFTESHLSTI
RINNDI"
complement(10812..11159)
/ gene="CpJ0276"
/ complement(10812..11159)
/ gene="CpJ0276"
/ codon_start=1
/ transl_table=11
/ product="CpJ91 hypothetical protein"
/ protein_id="BA98486.1"
/ db_xref="GI:8978650"
/ translation="MFLKRRKGGSSQVQNKRTASPIKHAHYLHNYLOELQIMARP
HDAIDAMQVFRDXYKKMSQAIIGFRDHLIVKVYNSLYALLKQTPQNDLMSLYOVA
SHVOIREQFLIG"

alignment_scores:
Quality: 947.50 Length: 910
Ratio: 1.771 Gaps: 32
Percent Similarity: 58.791 Percent Identity: 31.648

alignment_block:
US-09-677-752-4 x AP002546 ..
Align seg 1/1 to: AF002546 from: 1 to: 300650

85 ThrAspProLySGuaIaLeuPheLyGluLySAspLeuSerIleGI 101
||||| : : : : : ||||| : : : : :
200438 ACAATGCAATGCACTATCTAATCTAACAGGGAGTGTCTCAATCAC 200487
101 nasnPheArgPhe...LeuSerPheThrAspCysSerSerLySGluSerS 117
||| : : : : : ||||| : : : : :
200488 CAATGCAGCATCTCCGACACCTTAACCGCTTCCTTTAAAGAACTA 200537
117 erProSerIleIleHisGlnLySAsnGly...GlnLeuSerLeuArgAsn 132
: : : : : ||| : : : : : ||||| : : : : :
200538 CTGGGATCTTTCTTCTCCAAAGCCACGCTCAATTTCTCTCAAAAT 200587
133 : : : : : ||| : : : : : : : : : : : : : : : : : : : : :
200588 ATCGATCCGGGACCAACTACCTTTACCAATACACCTCCAAATTAAGCT 200637
135 rMeuSerPhe...CysArgAsnH 142
||||| : : : : : ||||| : : : : :
200638 TCTCTCTTTTACAGATTCTTCCTATTGTTCACATAATCAAAACCGAATG 200687
142 tsaIaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSerLeuGln 158
: : : : : ||| : : : : : ||||| : : : : :
200688 CTACCCAGACGAGACGAGCATCAAGTCCACAGAGCTTGTCTTATTCAG 200737
159 HisAsnTrpLeuPheThrAlaPheGluGluAsnSerSerLySGlyAsnGI 175
||||| : : : : : ||||| : : : : :
```

```
200738 TCGAAGTAT...AGTTGCTACTTTGGCCAAACTTTTCTATATGACATGG 200784
175 yGlyAlaIleGlnAlaGlnThrPheSerLeuSerArgAsnValSerProI 192
||||| : : : : : ||||| : : : : :
200785 AGCGGCCCTCCAAAGGACGCTCTATCAGTATATCGCTAAACCCCAAC...C 200831
192 leSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCysCys 208
: : : : : ||| : : : : : ||||| : : : : :
200832 TAACGTTTGGCAAAACAAAGCAACCAAAAGGGGTGGCCCTCTATTTC 200881
209 SerAsnLeuIleCysSerGlyAsn...ValAsnProLeuPheThrArgI 224
||| : : : : : ||||| : : : : :
200882 ACCGAGCGGATTCAATTAATACCTTAACCTACACATCATTTTCTCA 200931
224 yAsnSerAlaThrAsn... 229
||||| : : : : : ||||| : : : : :
200932 AATTAACCGCGGCAACAATATGGCGAGCATTTACAGGAAGTAGAGTT 200981
229 : : : : : ||| : : : : : ||||| : : : : :
200982 TTATTAGCAGCAACAAGCAATTAGCTTTATAACAATAGTGTGACCGCA 201031
230 : : : : : ||| : : : : : ||||| : : : : :
201032 ACCTCAGCTACAGCGGAGCATTTACTGTAGTAGT...ACATCAGC 201075
||||| : : : : : ||||| : : : : :
242 rGluLySGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAlaS 259
||| : : : : : ||||| : : : : :
201076 CCCCAACACGAGCTTAATCTATCAGACACGCGGAGCACTGACTATAG 201125
||||| : : : : : ||||| : : : : :
259 erAsnSerAlaLySGlyGlyGlyAlaIleThrAlaLySHisMetVal 275
: : : : : ||| : : : : : ||||| : : : : :
201126 GAATATACGCAATTAAGTAGTGGTGGCGGATTATATCTGACAAATCTAGT 201175
||||| : : : : : ||||| : : : : :
276 LeuArgTrpAsnGlyProValSerPheIleAsnAsnSer... 288
||| : : : : : ||||| : : : : :
201176 CTTCCTTCCTGAGAGACCTACCGCTTTTAAAAACACTCTCTATAGTAC 201225
289 : : : : : ||| : : : : : ||||| : : : : :
201226 TGCAGCTCCCTTAGAGAGCAATTCGATTGCTGATCTGATCTTTGA 201275
||||| : : : : : ||||| : : : : :
304 erIleLeuAlaGlyGlyGlySerValLeuPheGlnAsnAsnSer... 318
||| : : : : : ||||| : : : : :
201276 GTCTTCGCTCTTGTGTGAGACATCATCTTTGAAGGAACACAGTAGTC 201325
||||| : : : : : ||||| : : : : :
319 : : : : : ||| : : : : : ||||| : : : : :
201326 AAGGAGACTTCTTGAGATCAGACACTACCAAGAAATTTCTTAACATCG 201375
||| : : : : : ||||| : : : : :
334 : : : : : ||| : : : : : ||||| : : : : :
201376 AAMCACCAGTCTAAGATTTGTACAGCTGCGAGCCCTCAAGCAATACTA 201425
||| : : : : : ||||| : : : : :
349 leuLeuPhePheAspProIleValGlnGluSerSerSerLySGluSerPro 365
||| : : : : : ||||| : : : : :
201426 TCTACTTCTATGATCTCTAACA... 201448
||||| : : : : : ||||| : : : : :
366 leuProSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAl 382
||||| : : : : : ||||| : : : : :
201449 : : : : : ||| : : : : : ||||| : : : : :
382 aSerProLeuValIleGlnThrSerAlaAsnArgSerValIlePheSerS 399
||||| : : : : : ||||| : : : : :
201493 TCTGACCTTCGAGGGAATCTCGATATCAAGAACCATGATTTTCTG 201542
||||| : : : : : ||||| : : : : :
399 erGluArgGluSerGlyGlyGlyThr...ProAspAsnLeuThrSer 414
||| : : : : : ||||| : : : : :
201543 GAGGAGAGCTCTGGAAGCAGAAAGCTGCAAGCTGTAATCTCAATCT 201592
||||| : : : : : ||||| : : : : :
415 GlnLeuGlnInpProIleGlnLeuLySergLyArgLeuValLeuLySAs 431
||||| : : : : : ||||| : : : : :
201593 ACATTTACGACACTCTAATCTTTCGCGGAGGCAACTCTCTCTTAATATC 201642
||||| : : : : : ||||| : : : : :
```


415 GluLeuGlnGlnProIleGluLeuLysSerGlyArgLeuValLeuLysAs 431
 5135 ACAATTAGCAACCTTAACCTTGGCCGAGGCACTCTCTTAAATC 5184
 431 PARGAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeu 448
 5185 AGGAGTCACTAGTGTCTAAGTCTTTTGCATCTCCGGGCTCAACC 5234
 448 euIleMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeu 464
 5235 TCCCTATGATGCGAGGACCATTAAGAACCCGTGATGGATCATCATC 5284
 465 AlaThrLeuSerIleProLeuHisSerLeuAspThrGlnLysSerValTh 481
 5285 AATAATCTTGTTCATGTAGATTCCTTAACACAGACCAAGAGGCTAC 5334
 481 rIleHisAlaProAsnLeuSerIleGlnLysIlePheLeuSerAsnSer 497
 5335 GCTAAAGCAACACAGCAAGT...CAGACAGTCACTTATCTGATCGC 5381
 498GlyAspGluAsnPhyThrGluAsnValGluLeuLeu 509
 5382 TCTCTCTTAGATCCTCTGGAATGCTCTAGAAAGTCTC..... 5422
 510 SerLysGluGlnAsnAsp.....IleProLeuLeuThrLeuSerLys 523
 5423TCTTGAATTAACCTCAACTCTTCTCTCTCACTCTTACTTC 5466
 523 sGluGln...SerHisLeuHisLeuProAspGlyAsnLeuSerSer 537
 5467 TGAGACCCCGCGAATATTCATCACCAGAC.....TTAGCTCTGATC 5510
 538HisPheGlyThrGlnGlyAspThrPhe 547
 5511 CCCTAGAAAAAATCTATCCATTCATGGGATACCAAGGGAATGGGCAATTA 5560
 548 SerTrpLysAspSerAspGluGlyHisSerLeuLeu.....AsnTr 562
 5561 TCTTGGCAAGGATCTGCGCACTAATTCCAAGCAGGCACTTACTCTG 5610
 562 pThrProLysAsnTyValProHisProGluArgGlnSerThrLeuVal 579
 5611 GACAAAAACAGATACATCCGAATCTCGAGCGTGGAACTTACTGTTG 5660
 579 IAsnThrLeuTrpAsnThrTyTrSerAspMetGlnAlaValGlnSerMet 595
 5661 CTACACCCCTATGGGATCTTGTGTGATGTGCGCTCATACAAAGCTT 5710
 596 IleAsnThrIleAlaHisGlyAlaTyLeuPheGlyThrTrpGlySe 612
 5711 GTAGCCACTAAAGTACGCCAATCTCAGAAACTCGCGCATCTGCTGTA 5760
 612 rAlaValSerAsnLeuPheTyAlaHisAspSerSerGlyLysProIle 629
 5761 AGGATCTCGCAACTCTTC.....CATAAAGATGACAGCAAGAAATA 5804
 629 sPAsnTrpHisArgSerLeuGlyTyLeuPheGlyIleSerThrHis 645
 5805 AAGGTTTTCGCGACATAGTGCAGGTATGTTGGAGCGCTACACACA 5854
 646 SerLeuAspAspHisSerPheCysLeuAlaIleGlnLeuLeuGly 662
 5855 TTAGCTTCTGAATCTTATCTACCTCCAGCTTCCCAATTAATGGGAA 5904
 662 sSerSerAspSerPheIleThrSerThrGluThrSerTyIleAlaT 679
 5905 AGATAGATCACTTATATAATAAATAGAGCTTCTGCTATGACGCT 5954
 679 hrValGlnAlaGln...LeuAlaThr.....SerLeuMetLys 690
 5955 CTCTCATCTCCAGCATCTAGGACCTTGTCTTCTCAAGCTTGTACGC 6004

691IleSerAlaGlnAl 695
 6005 TACCTTCTGATCTGAAGTGAAGCAGCAGCTCTCTCTTATCTCTCAGAT 6054
 695 aCysTyAsnGluSerIleHisGluLeuLysThrLysTyArgSerPhe 712
 6055 CAGCTATATCTATAGTAAATACTATGAAACCTATACACCAACGAC 6104
 712 eArgGluGlyPheGlySerThrPheSerValAlaValSerGlyGluVal 728
 6105 CAAAG...GGAGACGCTCGTATATAGACGGTTCGCTGAGACTT 6151
 729 CysAlaSerIleProIleValSer...AsnGlySerGlyLeuPheSer 744
 6152 GCGAGCTCCTACACACACTGCTTAAAGCATGAGGCTCTCTCCACC 6201
 744 rPheSerIlePheSerLysLeuGlnGlyPheSerGlyThrGlnAspGly 761
 6202 GTATTTCCTTCTCATCAAGTAGAAGCTTGATCATACCAAGATAGCT 6251
 761 heGluGlu...SerSerGlyGluIleArgSerPheSerAlaSerSerPhe 776
 6252 TCAGAAAGACGTAAATACACTTGTAGCATCTTTCATACGGGATTTA 6301
 777 ArgAsnIleSerLeuProIleGlyIleThrPheGluLysLysSerGln 793
 6302 ATTAACGCTCTGCTGCTATTGGAATACCTTCGAGAGATTCGAGAAA 6351
 793 sThrArgThrTyThrTyThrPheLeuGlyAlaTyIleGlnAspLeuLys 810
 6352 CGAGCGTGGCTTACGAGAGCTACTGATCATCTAGTCCGATCTATC 6401
 810 rGAspValGluSerGlyProValValLeuLeuLysAsnAlaValSerTrp 826
 6402 GTAAGATCTCTGATGACGACGACCTCTCTTAATCAACAATTAACCTCG 6451
 827 AspAlaProMetAlaAsnLeuAspSerArgAla..... 837
 6452 AAACCTACAGACGAATCTCTCAAGCAACCTGTATCGGAGAGCAG 6501
 838TyMetPheArgLeuThrAsnGlnArgAlaLeuHisArgLeu 852
 6502 GATCTTTATCTCTCTCT.....CCAATCTTG 6530
 852 InThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSerTrp 868
 6531 AGGTACACAGTACCTATCTATGCAAAATTCGTGATCTTCACGCACTAC 6580
 869 SerLeuAspLeuGlyThrThrTyArgPhe 878
 6581 AATGACATCTTGGAGGTAACTTCCACTTC 6610
 seq_name: gb_ba2:AP002546
 seq_documentation_block:
 LOCUS AP002546 300650 bp DNA BCT 08-DEC-2000
 DEFINITION Chlamydothila pneumoniae J138 genomic DNA, complete sequence,
 section 2/4.
 ACCESSION AP002546 AB033782 AB033783 AB033784 AB033785 AB033800 AB033801
 AB033802 AB033803 AB033804 AB033805 AB033806 AB033807 AB033808
 AB033809 AB033810 AB033811 AB033812 AB033813 AB033814 AB033815
 AB036079 AB036080 AB036081 AB036082 AB036348 AB036349 BA000008
 VERSION AP002546.2 GI:10176692
 KEYWORDS
 SOURCE Chlamydothila pneumoniae J138
 ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydothila.
 REFERENCE 1 (sites)
 AUTHORS Shiraï,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,
 TITLE Comparison of whole genome sequences of Chlamydia pneumoniae J138
 from Japan and CW029 from USA
 JOURNAL Nucleic Acids Res. 28 (12), 2311-2314 (2000)

REFERENCE 1 (bases 1 to 17280)
AUTHORS Hjerno,K., Boesen,T., Daugaard,L., Knudsen,K., Madsen,A.,
Christiansen,G. and Birkelund,S.
TITLE Chlamydia proteins containing the GGA1-repeat belong to a subfamily
of autotransporting pathogenicity factors
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 17280)
Boesen,T.
REFERENCE Direct Submission
JOURNAL Submitted (21-JAN-1999) Boesen T., Department of Medical
Microbiology and Immunology, University of Aarhus, The Bartholin
Building, DK-8000 Aarhus, DENMARK
FEATURES
source location/Qualifiers
1..17280
/organism="Chlamydia pneumoniae"
/strain="VR310"
/db_xref="taxon:83358"
1..841
/gene="omp12"
<1..841
/gene="omp12"
/note="putative"
/codon_start=2
/transl_table=11
/product="outer membrane protein 12"
/protein_id="CA837067.1"
/db_xref="GI:4455882"
/db_xref="SPTREMBL:Q92504"
/translacion="EDNIRYRHNSGVALGATATPAEDQLTFAFCOLFARDRNHITG
KHNQDYGCALYFHTEGLEDIANFACATRAPWLYSEISQI1PLSPFAKESYLIHD
NMKRYTNDNSI1KGSMDARFADLCALPVI5YVLLKEVPEPVKQYIYAHOOD
FERRHAGCAPKSKSELIINEIPICTYFFHDSSEKQTYDLTLMYIILDATRRNPKCOTS
LIASDANMAYGTNLARQGSVRAAHFQVNPHEMTEFGFAFEVRSSRNNTNLGSK
FCF"
885..967
/gene="omp11"
885..3677
/gene="omp11"
885..3677
/note="putative"
/codon_start=1
/transl_table=11
/product="outer membrane protein 11"
/protein_id="CA837068.1"
/db_xref="GI:4455883"
/db_xref="SPTREMBL:Q92393"
/translacion="MK1PLKLLISSTLVTPILLISATYGAADASISPTSPDQAGST
FIPKSTADANGTNYVLSGVNYINDACKGTALICCCPTTGTGDLTFPGKGYSSFNIVD
ACSNAGAASSTADAKALITTCFSNLSYIAAPGTVVASKSTISSACALNITDNGTILF
SOWNSGEANNNGALITTKLISGNTSSITFTSNSAKKLGAIYSSAASISONTQL
VEMNNKGTGGGALGFEASSITQNSLFECSNTADAGKGAICYCEKGTGEPPLTI
SENKSLTFPENSSTVOGGAICAHGIDLASAGPTLFPNNRCGNTAAGGALIAIDSGS
LSISANOCQITFLGNTLITSAFTSRNALYLCSSAKITNLAAOCOSIYFVDPILASN
TGASVDLITINQDENSPLDYSGTIVFSGEKLSEDAKKAADNFTSILKQPLASQPL
ALKQVELDVNGFTQREGSTLLMOPTCKKADPEALISLKLIVDLSALGKNSYSIET
AGAKNTITLSPLEVDSSGNTESHTINQATFQPLVFTATATASDIITDALITLSPV
QPEPEHYGQGHMEATWADTSIAKSGTMTWVTGGYNPNERASVVPDLSLMSFTDIR
TLQOITSOANSIYQOGLMASGTANFHKDKSGTQAFRIKSYGIYVGSSEDFSEN
IFSVAFCQLEFGKDLFLIVENTSHNYLASILYOHRAFLCGLPMPSPGSLTDMKQDPL
LINAOLSYSTRKNDMTRYTSYPEAGCSNNSGALLEGISALYLPKRAPFPQCGFP
FLKFOAVYSROONFKESGAEARAFDDGDYLVNSIPGICRILEKISDEPKNPELSLAYI
GDVYRKNPFRSRTSLVSGASWTSLECNLAROAFLASAGSHLTLSPIVELSGEAAVELR
GSAHLYNDCGARYSF"
3827..3907
/gene="omp10"
3827..6613
/gene="omp10"
3827..6613
/gene="omp10"
/note="putative"
/codon_start=1
/transl_table=11

product="outer membrane protein"
/protein_id="CA837069.1"
/db_xref="GI:4455884"
/db_xref="SPTREMBL:Q92398"
/translacion="MKSSLHWFILISSLALPLSLNFSAAVAVINILGPTSPSGPT
VTPPAOTNAOCTIYNLGDVSTINAGSPETALTACSPKFTGNSISPOCHQOFLONI
DAGACTPNTNANKLLSPSGSYLSLIOTTRATTTGCAIKTCACSIO5NSCTFGQ
NFSNDNGALOGSSISLSLNPULTEKAKNAOKGALYSTGITTNNLNSAFSENT
AANGCAIYTEASSSEFISNKAISFTNNSTVATASAGAICYSTSPAPVLLISNGE
LNFINTAITSGAIYTDNLVLSGGPTLFKNNSAIDPAAPGALAIADSGLSLSA
LGDDTFEGNTVVKGASSQOTTRNSINGTNNAKIVOLRASOGNTIYFDPITTSIT
AALSDAI.NGPNPLDAGNPAYOCTIVFSGEKISEAARAEADNKSITQOPLTAGOLS
LKSGYTLVAKSFSGSPGSTILMDACTTLETADGITTNNLVNVDLSKETKATLAKTO
ASQVTLTSGSLSDVPSGNVYEDVSMNPOVFSCLTLADDPANIHITDILAADPLEKN
PIHMGVQGNWALSMOEDTATKSKAATLTMTKGYNPNERGTLVANTLMSFYDVS
IQOLVATYROSQETRGWCEGISENFHKDSTKINKGFRHISAGYVATTTIADSLN
ITAFRCOLEGKRDHIEINKRASAVAASLHLOHAKTLSSPILRYLPGSESPQVLPD
AOISYISKNMKTYTQAPKGGSSMYNDGCALELASSLPHALSHGCLFHAYPEFIK
VBA5YIHQDSFEKERTTLVRSDDSLDNLVNSIPICITTFERRSRNREARVETIYIAD
VYKPNDCCTTALLINNTSWKTTGTNLSMQAGIGRAGIFYAESPNLEVSINLSMEIRGS
SRSYNADLGKRFQF"
complement(5954..8060)
/note="ORF1"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CA837070.1"
/db_xref="GI:4455885"
/db_xref="SPTREMBL:Q92897"
/translacion="MSPPQDPEQCSFRDVDSGLFRPESILTRASDFEGRIVYEWKRY
EBAALRNLIKQTEAGLIFPTQGEFRPRYSMDPEFMGHCYDNRDSDNPEGYVLRD
KISVSRHPTIEHFEVKTPEKGNNAKQRTISPSQFHEHIFAPMLKTRKREYVNOE
LIDDIYFYROYIODLYAAGCRNLIQDLCAMRLDIRP5VYGVSDSHLOEILFEO
LMIHNLVNMKREDPELFVSLMVCRGDYQVQEFERRRSLDIEEPLFAKTVDSYHFWAL
DKYSGAGPLAYVSGEKHVCGLISNSHSCIEDDVAVYSRIYEASVITPLERLSLP
complement(8586..11372)
/gene="omp5"
complement(8586..11372)
/gene="omp5"
/codon_start=1
/transl_table=11
/product="outer membrane protein 5"
/protein_id="CA837071.1"
/db_xref="GI:4455886"
/db_xref="SPTREMBL:Q956P2"
/translacion="MKSQFSMLVLSSTIACFPSCSTVEATENIGPDSDFGSGTNG
TYTPNNTTGIDYTLTGQITLGNLDSALITKGCPSDTESESIFAGKCYSLSPILKS
SAECAAISVTTQKNLSLIGFSSLTFLIAASSYITTPSCGAYKCGDGLTPDNKGITLF
KODYCEBNGCAISTKNLSLKNSTGSISEFGNKSATGKKGCAICATGTVDTNNAPT
LEFSNNIAEANGAIAINSTGCTTGNTSILVESNSVYATANGGALSADGADVTISNOS
VYFSDNOAVANGAIAAYAKKTLASGGGGGISFSNNIYVGTAGNGAISILAAECSL
SAEADDTFNGNAIYATTPQTRNSIDIGSFIJTNLRAISGHSIFEYDPTTANTAA
DSTDTLPPKKAAGNSTDVSIGSIYVSGEKLSEDEKAVADNLTSTIKOPTYTLAAGLV
KGVLTLDKFCPTQFQTSQSVIMAGCTYKASTREEVLTGSLIPVDSLGGCKKVIYASA
ASKNVALSPPILLLDNOGNAVENHDLCTQDSEPVQLSLGATTTTDPVAPVTVAPT
HYGYQWMTWVDVDTASTPKTKTATLANTNIGLYLPNPRQPLVPSNLWGSFSDIOA
IOGVERSALTLCSDRGFWAANVAFLEKDGKGRKRRHKGGAIVAGIAGAACSENL
ISFACQULESGDKFLVAKNHDTAGAGYIOHIRECSGFIKLDKIPGSSHKPLV
LEGOLAYSXNDKLTCTYAPYEVGSMCNAMFMMLKGLSSISYCEYELHCPPTVAPYI
KNTLYIIRQDSEKGTGCRSPDNLPLSLIPDYKPEKSCDCNDFSDLTLSYVD
LIRNDPKCTTALIVISGASMETYANNIAROLOVYRAGSHATSPMEPVEVLQOFVFEVGS
SRIVNDLGKRFQF"
complement(11313..11372)
/gene="omp5"
11313..14321
/gene="omp4"
11313..14321
/gene="omp4"
11313..14321
/gene="omp4"
/codon_start=1
/transl_table=11

sig_peptide
gene
CDS

sig_peptide
gene
CDS

```
glimmer2: putative"
/codon_start=1
/translation=1
/product="conserved hypothetical protein"
/protein_id="AAF38162.1"
/db_xref="GI:7189232"
/translation="MEMSPFQEQECHDVVSGFLRPESLTARSDFEGRIVEOM
RYVEPAIRNLKOTFAGLLEPTDGERRYMDPDSQFHHMTRAPNLKTRKPYPT
LKDQISVSKHPIEFERVKTEKNAKAKOTIPSPKFFHMTIRAPNLKTRKPYPT
NOELIDIVFYROYIOLYVAGCRNLOLDDCAWCRLOLDIRAPNSGYSDHROLEIL
EOLFVHNLVMDREDEFLVSLHVCRGDQAFESRRAYDSIEPLFAKTDVSYHY
NALDQKYSAGAPLAYSGEKHVCGLTSSNSCLIEDRDAVVSRIEASVYPLRLSL
LSPOGASCEGDHMTFEEOKKTAIVKEIKELMG"
/complement(9785..12571)
/gene="CP0306"
/complement(9785..12571)
/gene="CP0306"
/translation="similar to GP14376731; identified by sequence
similarity: putative"
/codon_start=1
/translation=1
/product="polymorphic membrane protein G family"
/protein_id="AAF38163.1"
/db_xref="GI:7189233"
/translation="MKSSLHMFLLSSSLALPLSLNSAFAAVEINLGPNSFSGPT
VTPAOTRADCTIYNLTCDSITNAGSPFATFASCFKRTGNLSPOCHGPFLLONI
DAGACFTNTANKLTSFSGSYSLIOTTNATGTGAIKSTGACSIOSNYCTFGQ
NEFNDGALTEQSSISLSLNPULTEAKNATKGALYSTGITTNNLTNSAFSENT
AANNNGAITEASSFISNKAISFTINNSVATAGALYSTGISTAPKPVLLSDGE
LNFICNTATSCGAIYTDNLVLSGCPPLKNNSAIDVPAIGALAIADGSLISA
LGGDITFEQNTVVGKASSSQITRRNSINIGNINAKIYOLKASOGNTIYEDPTTSIT
AALSDALNNGPDIAGNPAYOCTIVPSGKLSAFAAEADNLSKSTIOQPLTLACQOLS
LNSGTVLAKFSQSGSTLLDADGTTLETAGCITNNLVLVNDSIKETKTKLTKTO
ASQTVLSSLSLVDPSGNVYEDVSWNNPOVSCUTLTDADDPAIITLADLPLEKN
PIHMGQNMALSMQEDTATKSKAATLTWTKTGYNPNRRRTLVANLTLDSEFVDRS
IOQLVATKVRQSEGTGWCEGTSNFEHSDTKINQFHHISAGYVAGATTLASDNL
IQAACQOLGKDRDHPINKNRASVAAASLHLALATVSPSLRYLPAGESRQPVLEF
AQISYISKNTKMTYTOAPKGSWSYNDGCALELASSLPALHSHEGLFHYAPPIK
VSASTIHDSFEKERTNIVRSFSDGLINVSPIGTFFERFERSPEASVEATVIVYAD
VTRKNDCTTALLINNTSKITGTINLSROAGIGRAGITAFSPNLEVTNLSMEINGS
SRSYNADLGKTFQ"
```

BASE COUNT 3580 a 2439 c 2890 g 3767 t

ORIGIN

alignment_scores:

Quality:	947.50	Length:	910
Ratio:	1.771	Gaps:	32
Percent Similarity:	58.791	Percent Identity:	31.648

alignment_block:

US-09-677-752-4 x AE002192/rev ..

Align seg 1/1 to reverse of: AE002192 from: 1 to: 12676

```
85 Thrspaproluylalaleuphelsylusglyaspheusleu101
|||||
12418 ACAATGCGAGATGGAATCTATATACTAACAGGGGATGTCTCAATCAC 12369
101 nasipheargphe...Leuserpetheraspyserserlysglusers 117
|||
12368 CAATGACAGATCTCGACAGCTCTAACCGCTTCCGCTTAAAGAACTA 12319
117 erpserleilehlsiglnylasngly...Glnleuserleuargasn 132
:::
12318 CTGGGAATCTTTCTTCCAGACGACGCTACCAATTTCTCTCAAAAT 12269
133 .....Asnglyse 135
12268 ATGATGCGGAGCGAAGCTGACTTACCAATACAGCTGCAAAATACCT 12219
135 rmetserphe.....Cysatrgasn 142
:::|||||
```

```
12218 TCCTCTCTTTCAAGATTCCTATTTGTCATTAATCAACACGCAATG 12169
142 ISAlaagluyseryglylyAlaIleSerAlaaspAlaPheSerLeuGln 158
:::
12168 CTACACAGACAGACGAGCCATCAAGTCACAGAGCTTCTTATTCAG 12119
159 HisanlyrleuphethrAlaPheGluGlnasnserserlysglansgl 175
|||||
12118 TCGAAGTACT...AGTTGCTACTTTGGCCAAACCTTTCTAATGCAATGG 12072
175 yglAlaIleGlnAlaGlnThrPheSerLeuserArgasValserProI 192
|||||
12071 AGCGCCCTCCCAAGCAGCTTATCAGCTATCCCTAAACCCCAAC...C 12025
192 IeserPheAlaIArgasNArgAlaaspLeuasnnglylyAlaIleCysCys 208
:::
12024 TAACTTTGCCAAACAAACAAACGCAAAAGGGGTGCCCTTATTC 11975
209 SerasnleuIleCysserlyasn...ValasnProleuphethrArgI 224
:::
11974 ACGGAGAGGATTAACATTAACATTAACATTAACATTAACATTTCTGA 11925
224 yAsnSerAlaThrAsn..... 229
11924 AAATACCGCGGCAACAAATGCGCGCCGACCATTTACACGGAAGCTAGAGTT 11875
229 ..... 229
11874 TTATTAACAGCAACAAACAAATTAAGCTTTAATCAATAGTGACCGCA 11825
230 .....GlyGlyAlaIleCysCysIleSeraspLeuasnThrse 242
11824 ACCACAGCTACAGGAGGAGCCATTTACTGTAGTACT...ACATAGC 11781
242 rGluylsGlySerLeuserleuAlaCysasnglylnThrleupheAlas 259
:::
11780 CCCCCAACCAAGCTTAACTATATCAGACAAACGCGAACTGCACTTATAG 11731
259 eAsnserserAlaIlysglylyAlaIleAlaIleGlnSerlysglyserleu 275
11730 GAAATACAGCAATTAAGTGTGTGGGGGATTTTATCTGACAAATCTAGTT 11681
276 leuArglyrasnglyProvalSerPheIleasnanser..... 288
11680 CTTCTCTTGAGGAGCCTACGCTTTTAAACAAACCTGCTATAGATAC 11631
289 ...AlaIlyIleGlylyAlaIleAlaIleGlnSerlysglyserleu 304
11630 TCCAGCTCCCTTAGAGAGGACCAATTCGATTCGACCTGATCTTGA 11581
304 erIleleuAlaIlyGlylyserValleupheGlnasnanser..... 318
11580 GTCTTTGCGCTCTGTGGAGACATCACTTTGAAGGAACACAGTATGC 11531
319 ...GlnArgThrseraspGlnlyLeuValArgasAlaIlelyrleu... 333
11530 AAAGAGCTTCTTGAGTCACAGCACACCTACAGAAATTTCTATTAACG 11481
334 GlnlyaspAlaIleleuserSerleuGlnAlaIArgasnlyasp...I 349
:::
11480 AAACACCAATCTTAAGATTTGACAGCTCGAGCCTTCACAGCAACTACTA 11431
349 IeLeuphePheaspProIleValGlnIlyserSerlysglyserPro 365
11430 TCTACTTATGATGCTATTAACA..... 11408
366 LeuproSerSerleuGlnAlaSerValThrserProthrProAlaThrAl 382
11407 .....ACTAGCATCTGACAGCTCTCTCAGATGCTCAAACTTAATAG 11364
382 aSerProleuValIleGlnThrSerAlaIArgasNArgSerValIlePheSers 399
11363 TCTGACCTTGACGAGATCTCGATATACAGAAACCATCTGATTTCTG 11314
```

```

1147 AAAGAGCTTTCTGAGTCAGACCCAGTACCGAATCTCTATTACATCGG 1196
      :::::::::::::: ::::::::::::::
334  GluLysAspAlaIleuSerSerLeuGluAlaArgAsnGlyasp... 1 349
      :::::::::::::: ::::::::::::::
1197 AAACACCAATGCTAGATTGATGACCTGCGAGCCTCTCAAGCAATACTA 1246
      :::::::::::::: ::::::::::::::
349  IeLeuPhePheAspProIleValGlnGluSerSerLeuGluSerPro 365
      :::::::::::::: ::::::::::::::
1247 TCTACTTCTATGATCTCTATVACA..... 1269
      :::::::::::::: ::::::::::::::
366  LeuProSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAl 382
      :::::::::::::: ::::::::::::::
1270 .....ACTAGATCACTGACAGCTCTCTCAGATGCTCTAACTTAATAGG 1313
      :::::::::::::: ::::::::::::::
382  AserProLeuValIleGlnThrSerAlaAsnArgSerValIlePheSerS 399
      :::::::::::::: ::::::::::::::
1314 TCTTGACCTTCAGAGGAATCTCGCATATCAAGAACCATCGTATTTTCG 1363
      :::::::::::::: ::::::::::::::
399  eGluArgLeuSerGluGluGlyThr...ProAspAsnLeuThrSer 414
      :::::::::::::: ::::::::::::::
1364 GAGACAAAGCTTCGGAGACAGCAAGCTCGAAGCTGATATCTCAAACT 1413
      :::::::::::::: ::::::::::::::
415  GlnLeuGlnProIleGluLeuLysSerGlyArgLeuValLeuLysAs 431
      :::::::::::::: ::::::::::::::
1414 ACAATTCACCAACCTCTACCTTCGCGAGGCAACTCTCTTTAATC 1463
      :::::::::::::: ::::::::::::::
431  pArgAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeu 448
      :::::::::::::: ::::::::::::::
1464 AGGAGTCACTAGTGTCTAAGTCCCTTTCGCAATCTCCGGCTCTACCC 1513
      :::::::::::::: ::::::::::::::
448  euIleMetIuAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeu 464
      :::::::::::::: ::::::::::::::
1514 TCTCATGCGATCCAGGACCATTTAGAAACCGCTGATGGATCACTATC 1563
      :::::::::::::: ::::::::::::::
465  AlaThrLeuSerIleProLeuHisSerLeuAspThrGluLysSerValTh 481
      :::::::::::::: ::::::::::::::
1564 AATAACTTGTCTCAATGATCTCTTAAAGAACCAAGCAAGGCTAC 1613
      :::::::::::::: ::::::::::::::
481  rIleHisAlaProAsnLeuSerIleGlnLysIlePheLeuSerAsnSer 497
      :::::::::::::: ::::::::::::::
1614 GCTAAAGCAACACAGCAAGT...CAGACAGTCACTTTATCTGATCCG 1660
      :::::::::::::: ::::::::::::::
498  .....GlyAspGluAsnPheArgLysAsnValGluLeuLeu 509
      :::::::::::::: ::::::::::::::
1661 TCTCTCTTAGATCCTTCTGCAAAATCTTACGACATGTC..... 1701
      :::::::::::::: ::::::::::::::
510  SerLysGluGlnAsnAsn.....IleProLeuLeuThrLeuSerLys 523
      :::::::::::::: ::::::::::::::
1702 .....TCTTGAATACCTCAAGCTTTTCTGTCTCACTCTTACTGC 1745
      :::::::::::::: ::::::::::::::
523  sGluGln...SerHisLeuHisLeuProAspGlyAsnLeuSerSer... 537
      :::::::::::::: ::::::::::::::
1746 TGAGGACCCCGCAATATTCACATCACAGAC.....TAGCTGCTGATC 1789
      :::::::::::::: ::::::::::::::
538  .....HisPheGlyThrGlnGlyAspThrPhe 547
      :::::::::::::: ::::::::::::::
1790 CCTAGAAAAAATCTATCTCATTTGGGATACCAAGGAAATGGGCATT 1839
      :::::::::::::: ::::::::::::::
548  SerTrpLysAspSerAspGluGlyHisSerLeuIleAla.....AsnTr 562
      :::::::::::::: ::::::::::::::
1840 TCTTGCGCAAGAGTACTGCGCAATATCCAAACAGCAGACTCTTACTCG 1889
      :::::::::::::: ::::::::::::::
562  pThrProLysAsnTyuValProHisProGluValGlnSerThrLeuValA 579
      :::::::::::::: ::::::::::::::
1890 GACAAAAACAGATACCAATCCGATCTCGAGCTCGTGGAACTTAGTGG 1939
      :::::::::::::: ::::::::::::::
579  IaaenThrLeuTrpAsnThrTrpSerAspMetGlnAlaValGlnSerMet 595
      :::::::::::::: ::::::::::::::
1940 CTAAACGCTTGGGATCTCTTGTGTGTGTGGCTCATACAAACGCTT 1989
      :::::::::::::: ::::::::::::::
596  IleasnThrIleAlaHisGlyAlaValArgLeuPheGlyThrTrpGlySe 612
      :::::::::::::: ::::::::::::::

```

```

1990 GTACCCACTAAAGTACCCCAATCTCAAGAAACTGCGGCATCTGCTGA 2039
      :::::::::::::: ::::::::::::::
612  rAlaValSerAsnLeuPheArgAlaHisAspSerSerGlyLysProIleA 629
      :::::::::::::: ::::::::::::::
2040 AGGATCTCGAATCTTTC.....CATAAAGATAGCAGCAAGATAATA 2083
      :::::::::::::: ::::::::::::::
629  sPasnTrpHisHisArgSerLeuGlyThrLeuPheGlyIleSerThrHis 645
      :::::::::::::: ::::::::::::::
2084 AAGCTTTGCCACATAGATGACAGTTATGTTGTAGGACCCATCAACA 2133
      :::::::::::::: ::::::::::::::
646  SerLeuAspAsnHisSerPheCysLeuAlaIleGlnLeuLeuGly 662
      :::::::::::::: ::::::::::::::
2134 TTAGCTTGTGATTAATCTTATCACTGACGCTGCGCAATTTATGCGGA 2183
      :::::::::::::: ::::::::::::::
662  sSerSerAspSerPheIleThrSerThrGluThrThrSerTrpIleAla 679
      :::::::::::::: ::::::::::::::
2184 AGATAGATATCACTTTAAATAAATAAGACCTTCGCTATGACGCTT 2233
      :::::::::::::: ::::::::::::::
679  hValGlnAlaGln...LeuAlaThr.....SerLeuMetLys 690
      :::::::::::::: ::::::::::::::
2234 CTCTGCATCTCCAGCATCTAGCAGCTTGTCTTCSAAGCTTGTACGC 2283
      :::::::::::::: ::::::::::::::
691  .....IleSerAlaGlnAl 695
      :::::::::::::: ::::::::::::::
2284 TACCTTCTGATCTGAAAGTACAGCAGCTGCTCTTGTGATGCTCAGAT 2333
      :::::::::::::: ::::::::::::::
695  acyTrpArgAsnGluSerIleHisGluLeuLysThrLysTrpArgSerPhe 712
      :::::::::::::: ::::::::::::::
2334 CAGCTATATCTATGATAAAATCATATGAAACCTATTAACACCAGAAC 2383
      :::::::::::::: ::::::::::::::
712  eLysGluGlyPheGlySerTrpHisSerValAlaValSerGlyValAl 728
      :::::::::::::: ::::::::::::::
2384 CAAG...CGACAGAGCTCGCTGTAAATGACGCTTGCCCTGTGAAGCT 2430
      :::::::::::::: ::::::::::::::
729  CysAlaSerIleProIleValSer...AsnGlySerGlyLeuPheSerSe 744
      :::::::::::::: ::::::::::::::
2431 GCGAGCTCCCTACACACAGCTCTTAAACCAAGAGGCTGTCTCACGCC 2480
      :::::::::::::: ::::::::::::::
744  rPheSerIlePheSerLysLeuGlnGlyPheSerGlyThrGlnAspGlyP 761
      :::::::::::::: ::::::::::::::
2481 GATTTTCTCTTATCATCAAGTAGAAGCTTCGTATACACCAAGATAGCT 2530
      :::::::::::::: ::::::::::::::
761  heGluGlu...SerSerGlyGluIleArgSerPheSerAlaSerPhe 776
      :::::::::::::: ::::::::::::::
2531 TCAAGACAGTATATCTACTCTGCTGACGATCTTTCCGATGCGGTGATTA 2580
      :::::::::::::: ::::::::::::::
777  ArgAsnIleSerLeuProIleGlyIleThrPheCysLysLysSerGln 793
      :::::::::::::: ::::::::::::::
2581 ATTAGCTCTCTGTGCTATGGAATTACTCTTGAGAGATCTCGAGADA 2630
      :::::::::::::: ::::::::::::::
793  sThrArgThrTrpTrpTrpPheLeuGluAlaValArgIleGlnAspLeu 810
      :::::::::::::: ::::::::::::::
2631 CGAGCTGCTCTTACCAAGCTCTGCTGATCTGATCTGATCTGATCTGAT 2680
      :::::::::::::: ::::::::::::::
810  rGAspValGluSerGlyProValValLeuLeuLysAsnAlaValSerTr 826
      :::::::::::::: ::::::::::::::
2681 GTAAGAACTCTGCTGACAGCAGCTCTCTCAATCAAACTACTCTGCTG 2730
      :::::::::::::: ::::::::::::::
827  AspAlaProMetAlaAsnLeuAspSerArgAla..... 837
      :::::::::::::: ::::::::::::::
2731 AAAACTACAGAAAGATCTCTGACAGCAAGCTGTTGTCGGAAGCAGG 2780
      :::::::::::::: ::::::::::::::
838  .....TyrMetPheArgLeuThrAsnGlnAlaValLeuHisArgLeu 852
      :::::::::::::: ::::::::::::::
2781 GATCTTTTATGCGCTTCTCT.....CMAATCTGTC 2809
      :::::::::::::: ::::::::::::::
852  IthrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSerTrp 868
      :::::::::::::: ::::::::::::::
2810 AGTCACAAGATCAATCTTATGGAATTTGCTGATCTTCAAGCAGCTTAC 2859
      :::::::::::::: ::::::::::::::
869  SerLeuAspLeuGlyThrThrTrpArgPhe 878
      :::::::::::::: ::::::::::::::
2860 AATGACATCTTGGAGTAGTACTTCACTTC 2889
      :::::::::::::: ::::::::::::::

```

```

483  ISAIAProAsnLeuSerIleGluLysIlePheLeuSerAsnSerGlyAsp 499
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
2293  AATCTCCAACTCTTCTATCCAAAAATCTTCTCTTAATTCAGACCAT 2244
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
500  GluAsnPheTyrGluAsnValGluLeuLeuSerGlyGluGlnAsnAsnI 516
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
2243  GAAATTTTATGAGATGTAGACCTATTAAAGCAAGATCAAAAAAGCAT 2194
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
516  eProLeuLeuThrLeuSerGlyGluInSerHisLeuHisLeuProAsp 533
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
2193  CCCCTCTCTCTCTCCAAAAGGCTGCCCAATCCAGATCTCTTCGATG 2144
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
533  LysAsnLeuSerSerHisPheGlyTyrGlnGlyAspTrpThrPheSerTrp 549
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
2143  GGAATCTTCTCTCTACCTTCTGATATCAAGGGCAATGCAATTTTCTTGG 2094
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
550  LysAspSerAspGluGlyHisSerLeuIleAlaAsnTrpThrProLysAs 566
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
2093  CAACATCTCCGATCAGAGAAACCTTAGTTGCTTAAGTGGACAGCTAATAG 2044
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
566  nTyrValProHisProGluArgGlnSerThrLeuValAlaAsnThrLeu 583
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
2043  CTACATACCAACCCCAACCCCAATCCCAATCCCAATCCCAATCCCTT 1994
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
583  rPAsnThrTyrSerAspMetGlnAlaValGlnSerMetIleAsnThrIle 599
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1993  GGAATTCAGTACTCTGACATGCAACCTGACATGATGATTAATACACACT 1944
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
600  AlaHisGlyGlyAlaTyrLeuPheGlyThrTrpGlySerAlaValAsn 616
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1943  GCTCAGAGAGAGGCTATCTATTGGAAACCTGGGGATCTGCCGTTTCTAA 1894
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
616  nLeuPheTyrAlaHisAspSerSerGlyLysProIleAspAsnTrpHis 633
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1893  TCTATTCTATTCACATGCAACCTGAGAAATTCACAGACACTCGCAAC 1844
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
633  iAspSerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAsp 649
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1843  ATGAGAGCTTGCTTACCTCTTCTGATATACAGCAGCATAGCTACAGC 1794
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
650  HisSerPheCysLeuAlaAlaGlyGluLeuGluGlyLysSerSerAspSe 666
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1793  CACGCTTCTCTCTAGCTGCTGAGCAATATTATCGGAAATCTTCAGATTC 1744
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
666  rPheIleHisSerThrGluThrThrSerTyrIleAlaIleThrValGlnAla 683
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1743  TTTTGTGACCTCTCGCAGACACCTCTTATATACACACTATACAAACAC 1694
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
683  InLeuAlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrAsnGlu 699
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1693  AAATAGCTACTCTTTAATTAATAATCTCTGCACAGGCTGCTATTAATGAA 1644
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
700  SerIleHisGluLeuLysThrLysTyrArgSerPheSerLysGluGlyPh 716
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1643  AGCATCCAGCAACTAAACAAATACCGCTCTTCTTAAGAAAGCTTT 1594
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
716  eGlySerTrpHisSerValAlaValSerGlyGluValCysAlaSerIleP 733
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1593  CGGTGCATGCGACAGCGCTCGCGATCCGAGAGATCCGCTCATCCATC 1544
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
733  rIleValSerAsnGlySerGlyLeuPheSerSerPheSerIlePheSer 749
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1543  CTATTGTATCCAAATGATCTGACCTTTTACTTCTCTCTCTCTTCTTCT 1494
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
750  LysLeuGlnGlyPheSerGlyThrGlnAspGlyPheGluGluGlnSerG 766
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1493  AAACATCAAGAGATTTTCAAGAAACAGATGCTTTGAAGAAAGCGGG 1444
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
766  yGluIleArgSerPheSerAlaSerPheArgAsnIleSerLeuProI 783
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1443  TGAACCTCGGGCTTTGCTGATAGCTCTTACATTAATATTTCGCTCTCT 1394
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
783  IeGlyIleThrPheGluLysLysSerGlnLysThrArgThrTyrTyr 799

```

```

1393  TAGGAATAGCATTCGAGAAATCTCAAAAACCTCAAACTACTACCAT 1344
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
800  PheLeuGlyAlaTyrIleGlnAspLeuLysArgAspValGluSerGlyPr 816
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1343  TTCTTAGACCTCATCATTCAGATCTAAACGCTGTGTAGAAATCGGGCC 1294
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
816  oValValLeuLeuLysAsnAlaValSerTrpAspAlaProMetAlaAsn 833
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1293  CGTAACCTCTTCAAAAATTCAGTACTTGGAGTCTCTATGGCAATC 1244
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
833  eAspSerArgAlaTyrMetPheArgLeuThrAsnGlnArgAlaLeuHis 849
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1243  TAGATTGCGAGCATGATGATTCGCCCTCAGAAATCAAAAGCTTTTAC 1194
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
850  ArgLeuGlnThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerH 866
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1193  AGATTTCAGACACTTGTGAATATGCTTATATGCTAGCTGGCGAGACTA 1144
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
866  sSerTyrSerLeuAspLeuGlyThrThrTyrArgPhe 878
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1143  TAGTTACTCTTGGATTTGGAACTACTTACAGATTC 1107
      ::::::::::::::::::::::::::::::::::::::::::::::::::::

seq_name: gb_bai:AE001628

seq_documentation_block:
LOCUS      AE001628      10757 bp      DNA      BCT      01-DEC-2000
DEFINITION Chlamydia pneumoniae section 44 of 103 of the complete genome.
ACCESSION  AE001628 AE001363
VERSION    AE001628.1 GI:4376730
KEYWORDS
SOURCE
ORGANISM   Chlamydia pneumoniae CML029.
            Chlamydia pneumoniae CML029
            Bacteria: Chlamydiales: Chlamydiaceae: Chlamydia.
REFERENCE
AUTHORS    Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R. W.,
            Olinger, L., Grimwood, J., Davis, R. W. and Stephens, R. S.
            Comparative genomes of Chlamydia pneumoniae and C. trachomatis
            Nat. Genet. 21 (4), 385-389 (1999)
JOURNAL    Nat. Genet. 21 (4), 385-389 (1999)
MEDLINE    99206606
PUBMED     10192388
REFERENCE  2 (bases 1 to 10757)
AUTHORS    Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,
            Grimwood, J., Davis, R. W. and Stephens, R. S.
            Direct Submission
            Submitted (01-DEC-1998) Program in Infectious Diseases, University
            of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
            Location/Qualifiers
            1..10757
            /organism="Chlamydia pneumoniae CML029"
            /strain="CML029"
            /db_xref="taxon:115713"
            106..2892
            /gene="pmp_9"
            /note="Cpn0447"
            106..2892
            /gene="pmp_9"
            /codon_start=1
            /transl_table=11
            /product="Polymorphic Outer Membrane Protein G/I Family"
            /protein_id="AAB18591.1"
            /db_xref="GI:4376731"
            /translation="MKSSLHWLFISSSLALPLSLNFSAAFAVEINLQPTNFSFGPT
            YTPPAQCTFNADGTYINLTGDSVITNAGSFTALJACFEKTTNLSFGQHGQVFLQNI
            DAGANCTFNNTANKLILSFGSFSYLSLIDTNAATGTGAIKSTGACSIOASNVCYFGQ
            NFSNDGALOGSSISLISLPNLTAKNKATOGKGLYSTGGITINLTNLSAASENT
            AANNNGAIYTEASSFSISNKAISFINNVATATSAAGAIYCSSTAPKPVLTLSNGE
            LNFIGNTAITSGAIYTDNLVSSGGPTLFTKNNASAIIDPAAPGAIATADGSLSLA
            LGGDITPEENTVYKKAASSQTTTRNSINGNTNNAKIVOLRAQGNITFYIPDITTSIT
            AALSDALNNGDLGNPAVQCTIYFSGKLSAEAEADNLKSTIQDPLTLAGOLS
            LKSGVTILAKSFSQSPGSLTLDADAGTTLETAGIT11NNLVWVDSLKEFKKATLTAQ
            ASQVTILSGSLSLVPSGNVYEDVSWNNQVFSCLTLEADPANHIHIDLAADPLEKN
            PIMHGYGNWALMSWEDTATKSKATLTWTKTGYNPNERGRTLVANLTMSFVDVRS

```

```
/gene="TC0265"  
241..495  
/gene="TC0265"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
/codon_start=1  
/transl_table=11  
/product="hypothetical protein"  
/protein_id="AAF39134.1"  
/db_xref="GI:7190306"  
/translation="MIIVGALRDPWLYFFLDPFTHVEKSLSTRNNKKGSLPFI  
PWGHOSCLFPHKVKVSMFVFPYKVGVGFVQRDLHLKK"  
426..1016  
/gene="TC0266"  
426..1016  
/gene="TC0266"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
/codon_start=1  
/transl_table=11  
/product="hypothetical protein"  
/protein_id="AAF39135.1"  
/db_xref="GI:7190307"  
/translation="MVCPLGGGLLRSEAGSSSEEVNODPLLDATISYAVRCLAVAK  
HTESATYHONKRCCKVCGMKVKRFLCCCCAEKWEKCTGICGLMOMETIRLHG  
IITGTAVHLSCVCEMMLAGFTLSFGDQRQINAAIEMSSITMRLMLDSSDGKVDG  
LSPAPQEDLEQYREKVGIGNLADAVALLHFCNR"  
complement(1104..3707)  
/gene="TC0267"  
complement(1104..3707)  
/gene="TC0267"  
/note="similar to GP:4376270; identified by sequence  
similarity; putative"  
/codon_start=1  
/transl_table=11  
/product="polymorphic membrane protein G family"  
/protein_id="AAF39136.1"  
/db_xref="GI:7190308"  
/translation="MRPDVNLCCCLCATLISPTAILFGDADLKSALITKNPSIVCT  
FLEDCTMENEPSPALLSHARQDDPLYIGNTNMFVNLSHPSTNEERFLKEKGLSTOD  
PFLSTDCSSSTEDSPILIHNGQLFRNNMGMSFYRNHSEGSGLSTDLALFH  
NLPFTFENSSAKNGAIOAQTLSLRNWSLSFSKRNKRNALNGCAICGONLCSGN  
NLPFTFNALNGCAICINEONLSKGLSLAYNOFTLFCNSKAEKGCALYTKRMV  
LRHNGVSVVNSAKLGAIAIOSGSLSLIAGGGSVLFONNSCHSDGVANALYL  
EKNALLSLEARRGDIIEFDPIVQEVVSEFTSALTPLRIQTNNRAVIESSENL  
KEKTEANLISKIOQPIELDSGLVLDKRVILSAPLSOAPQALLVMADVGTSLSD  
LKLTSLIPLHSIDTENSYSIOSPTLSIOKIFLSENHEHYENVLLSKDODIPL  
SLPKGLPHDPLDQNLSSHFGOGDMNFSQMSDQRETLVAMTANSYLPHPEQSAL  
VANTLNNITSDMOAVOSMINTTAQGCATLFGTWGSAVSNLFSHGNSGASTDMWKRK  
LGYLFGISTHSLDDHSFCLAAQQLFGKSSDSFVTSADTTSYIAAIDTQIATSLIKISA  
OACYNESIHETLKRYRSFGKGFAMHSAVSGEIASIPIVNSGGLSFSEFISKL  
OGFSGKQDFEESRGEARAFADSFTNISLAPVIAFEKKSOKTRNYHFLGAYIDLK  
RCVESGPVTLTKNSVTWAPMANLDSRAMMFLITNRALHREQTLYNMSYMLGQSYS  
YSIDLCTYRF"  
3956..5800  
/gene="TC0268"  
3956..5800  
/gene="TC0268"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
/codon_start=1  
/transl_table=11  
/product="hypothetical protein"  
/protein_id="AAF39137.1"  
/db_xref="GI:7190309"  
/translation="MFYFLGMFVWGIKGVGSGSDSYPIPSHNGDESEKNSDSTSS  
KYNKATSSLOCAPSTNDENSVSPYSDVDTLIDSGESSRHYIKSILTEHAHES  
SVBGACHSSROIICRIQACIGLARARVEKVRVTQSTIFPQACAGIORTGKATIKIS  
SASRGRLMTDTPWRYRLVRNPPMDGLFAKLDADAEADMAATKCYVSNLEKRGAD  
RETIENCLQWAKMEKRRARDLRMGAAKAKFLRDPFGKSDPKYKGTLPGEYTVONTMY  
DGQVSKLSEVDTGLDMEKLSDAVLSINIGLARARVINOISIPOLSEIERFVKL  
ESACDEARASLKEACWIKGKEPNKAORAFRRVEESRNLEISFGSGSARLSARY  
SQGLAAGCAIRRCFDCRCKKCVSLKKDLSEELNLAEELLIRFDMEGIERDPGNNTI
```

```
PHVENNRIGVPVIEEGGAETHIETMPOQESBOYEVWMDGLSEERDFAVSQOHYOV  
PRRSUNTEPRPREYDPRNSARSTYDVPRPDPONEVEKHVYTGMRSSYACFVAG  
MKNRYISQPEQIPNSEOYEQLFQELINDGDOIIDELKIMNEELDNO"  
complement(5923..6192)  
/gene="TC0269"  
complement(5923..6192)  
/gene="TC0269"  
/note="conserved hypothetical protein; identified by  
Glimmer2; putative"  
/codon_start=1  
/transl_table=11  
/product="conserved hypothetical protein"  
/protein_id="AAF39138.1"  
/db_xref="GI:7190310"  
/translation="MLCKVGRSLSLIIVGLAINAGILCATGYRVNIVARLLBEGTML  
TOAAVYVAGIACVIAFLNFECCKFKRHSDCCSSXGCGHHHTRD"  
6393..6695  
/gene="TC0270"  
6393..6695  
/gene="TC0270"  
/note="glutamyl-tRNA(Gln) amidotransferase subunit C,  
putative; identified by match to TIGR protein family HMW  
TIGR00135"  
/codon_start=1  
/transl_table=11  
/product="glutamyl-tRNA(Gln) amidotransferase subunit C,  
putative"  
/protein_id="AAF73542.1"  
/db_xref="GI:8163182"  
/translation="MTQSYVNNKEELISLAKNAALELEAHVEEFTSMNNVATLMOEV  
VALDISDITLEATVHHFVCPEDLRDMVTSDFTRREFLANVSVSLGGLKVPYVTK"  
6707..8182  
/gene="TC0271"  
6707..8182  
/gene="TC0271"  
/note="similar to GB:A009126; identified by sequence  
similarity; putative"  
/codon_start=1  
/transl_table=11  
/product="glutamyl-tRNA(Gln) amidotransferase subunit A"  
/protein_id="AAF39139.1"  
/db_xref="GI:7190311"  
/translation="MKRSALERDAVVVREISVATITFPFHRIESTDEQIGSLSL  
CEKRALLRASRIDKLAKGDPICITAGIPIGVKNDHTITGVTTTCAKMLNFPVAFD  
AVVRRITEMEDDILLEGKLMDEFAMGSTRYSAPFHTNPMOLERYPGSSGSAAY  
SARFCPIAGSGTGSIRPPAFGCVGKPSYGAVSRYGLVAFGSLDOIPLTVTV  
EDVALAMDQFAGGDIIDATTROPEFRGTFESQALSLEPKLIGVPRGLDLOEDKENE  
FPAALAMEROSRIIDISYLKHNVPYVITYASAAANLARPQVRRGHCACADN  
MOBTARSKRKEGKQEVTRILLGNVISAERONTFFYKGTAVRATLLEAFQSAFEC  
DVIAMPVCASPAIRDTDVLDPSLYLQDIYVAVNLAVLPALISVPSGLSKELPLGVO  
FIGKRGADQOICQVGSFQEHSDIKOLYKAVNGLFDGME"  
8184..9650  
/gene="TC0272"  
8184..9650  
/gene="TC0272"  
/note="similar to PID:1653395; identified by sequence  
similarity; putative"  
/codon_start=1  
/transl_table=11  
/product="glutamyl-tRNA(Gln) amidotransferase subunit B"  
/protein_id="AAF39140.1"  
/db_xref="GI:7190312"  
/translation="MGIAHTEWESVIGLEVHVELNTASKLSPARNHFGDEPNINSP  
VCTGMGALPVLNKDAVRKAVLFGCAVEDVALSFSDKSYFYDPSPNFOITQYEH  
PIYRGCVRAIVEGEKTEFELAOHTLEDAGMLKHGDFAGVDYNAAGVPLLEIYSKP  
CMSEADVAAYANALVSLIYGISNNEECSVRPDVNIASRPKSGSLRKAKVELKN  
NMSFTMAOALAEKHROJLEBILSHNEDNEPKVYPAATIRMPPEKKTYLMLKEAE  
DVMYVEPDLVPIQITESTIDEVRQTLPELPSKTKRYITFDLLEADLMLIVSDRHT  
AHFEETATDSKNYSRALSMITVEERAGKATGTLPTFGILPEWVAQDVNIDGVI  
TAKIEAIVKSNVSGSESPEDILRRHPSLLPMTDHALRAIVKEVVAQDNASVEDYK  
NKAARLGLVQCIKMRTEGKAPRKVNELLAAAMD"  
complement(9731..10834)  
/gene="TC0273"
```

gene

```
RSVIFSSERLSEEEKTPDNLTSLQLOPIELKSGRLKIDRAVLASPLSDPOALLIM
EAGTSLSKTSDDLKIALTLPLSLDTREKSVTHAPNLSTOKITLSNSGDNFENIEL
LSKEONNIPLLTLSKESQSLHLPDGNLSHFQGDWTFESWKSDSDSHLIAMTRKN
YVHPEROSTLVANTLANTYSDMQAVOSMINTLHGAGVLEGTWGSVSNLFLYAHDS
GRPIDMWHRSLSGLYLFISTHSLDSDHSECLAGOLGKSSDSFITTSTETSYIATVOA
OLATPLMKISAQACYNESIHELKTKYRSFSEKSGSHSVASVGECAISIPIVNSG
LFSSFSIFSKLQGFSGODFPESSSGEIRFSASSPRNLSLPMGITEPEKKSOKTRANY
VFGLAVIIDLKRDVESGPVLLKNAVSWDAPMANLDORAIMFRLITNORALHRLQTLIN
VSIVLRGSHSHSYSLDLGTYRPF"
3725..5500
/gene="Ct875"
3725..5500
/gene="Ct875"
/transl_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC68473.1"
/db_xref="GI:3329351"
/translation="MSTRGVGNGNRSRIPSHNGDGNRRSNTQKNKKEVDNCSLVS
SRSNENRESPIYAVDVSMLESTPTSGETTRASKGVSRRORGLVRAADKRRAYOCA
WSVSTRRSATRAEAGSSSRTRARGASSGREYSPSAARGLRLMFTDFMTRVLRQT
SPMAGVGNLDVNEARLMAATSECAHLEANKLAGDGAAREIAKRMEOVRDLQ
DKGAARLNDPLGRRTPNYOSKNRGEXTVGNMGPVAVANONVDGFLMSNL
SDVYLSREIQGLAARATLESMPLENERFRRLQETCOARTIEIESGWTRESAS
RMEDEAGPSPRAOQAFQSFVNECNSTIEFSFGSEHYRVLCAVRSRDLAAGAIR
CFSCCKGSTRHVARPDLSPGASLAETLAPADMGIERGADCTYDILPYDDBRCY
PSIEGEGSDSIYEIMPIYEVMDMLETRSFAYQOGHYODPARSDYDLPASVDLP
RSPPTPLPRYOLNDMDVEAGFREAVYASFAVMYVVTQOERLIPNSQVEGIL
RDMLTNSQTFRDLRMNRNREVRE"
```

BASE COUNT 1664 a 1030 c 1382 g 1525 t
ORIGIN

alignment_scores:
Quality: 4509.00 Length: 878
Ratio: 5.147 Gaps: 0
Percent Similarity: 99.772 Percent Identity: 99.544

alignment_block:
US-09-677-752-4 x AE001361/rev ..

Align seq 1/1 to reverse of: AE001361 from: 1 to: 5601

```
1 MetArgProAspHisMetAspHisCysCysLeuCysAlaIleLeuSe 17
|||||
3451 ATGGGACCTGATCATGAACTTCTGTTGCTATGCTGCTATTTGTC 3402
|||||
17 rSerThrAlaValLeuPheGlyGlnAspProLeuGlyGlnThrAlaLeuL 34
|||||
3401 ATCCACACCGCTCTCTTGGCCGACGATCCTTAGGTGTAACCGCCCTCC 3352
|||||
34 eutThrLysAsnProAsnHisValAlaCysThrPhePheGlnAspCysThr 50
|||||
3351 TCCTAAATAATCCTAATCATCTCTCTCATATTTTTTCAGACACTCAAC 3302
|||||
51 MetCysSerLeuPheProAlaLeuCysAlaHisAlaSerGlnAspAspPr 67
|||||
3301 ATGGAGACCTCTTCTCTGCTGCTGCTCATGCATCAACAAGATGATGC 3252
|||||
67 OleuTyrValLeuGlyAsnSerTyrCysTrpPheValSerIysLeuHisI 84
|||||
3251 TTTCTATCTACTTGAATAATCTCTACTGTTGGTTCCTATCTAAACCTCA 3202
|||||
84 IetHisAspProLysGlnAlaLeuPheLysGlnLysGlnAspLeuSerIle 100
|||||
3201 TCACGACGACCCCAAGACGCTCTTTTAAAGAAAAGAGATCTTTCATTT 3152
|||||
101 GlnAsnPheArgPheLeuSerPheThrAspCysSerSerLysGlnSerSe 117
|||||
3151 CAAATTTTTCGCTTCTCTTCACAGATGCTCTTCCAAAGAAAGCTC 3102
|||||
117 rProSerIleIleHisGlnLysAsnGlyClnLeuSerLeuArgAsnArg 134
|||||
```

```
3101 TCCTTCTATATTATTCATCAAAAGATGCTCACTTATCCTTGGCAATATG 3052
134 IySerMetSerPheCysArgAsnHisAlaGlyGlySerGlyValAlaIle 150
|||||
3051 GTAGATAGAGTTTCTGTGCAATCATGCTGAGAGGCTTCGGAGAGACCTC 3002
|||||
151 SerAlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPheG 167
|||||
3001 TCTGGGATGCTTCTTCTCTACACACACACTATCTTTCACAGCTTTTGA 2952
|||||
167 uGlnAsnSerSerLysGlyAsnGlyValAlaIleGlnAlaGlnThrPheS 184
|||||
2951 ACAGAAATCTTCTTAAGGAAATGCGGAGCCATTCAGCTCAACCTTCCT 2902
|||||
184 eLeuSerArgAsnValSerProIleSerPheAlaArgAsnArgAlaAsp 200
|||||
2901 CTTTATCTAGAAATGTCGCTATTTCTTGCCCGGATATGTCGGAT 2852
|||||
201 LeuAsnGlyGlyAlaIleCysCysSerAsnLeuIleCysSerGlyAsnVa 217
|||||
2851 TTAATGGCGGCGCTATTGCTGATGATCTTATTGTTTCAGGGAATGT 2802
|||||
217 IAsnProLeuPhePheThrGlyAsnSerAlaThrAsnGlyGlyAlaIleC 234
|||||
2801 AAACCTCTCTTTTTCACATGGAACCTCCGCCAGCAATGAGCGCGCTATT 2752
|||||
234 ySerLysIleSerAspLeuAsnThrSerGlyLysGlySerLeuSerLeuAl 250
|||||
2751 GTTGTATCAGCGATCTTAACACCTCGAANAAGGCTCTCTCTCTTGGCT 2702
|||||
251 CysAsnGlnGlyThrLeuPheAlaSerAsnSerAlaLysGlnLysGlyG 267
|||||
2701 TGTAAACCAAGAAAGCGTATTTGCAACCAATTTGCTAAAGAAAAGCGCG 2652
|||||
267 yAlaIleTyrAlaLysHisMetValLeuArgIlyrAsnGlyProValSerP 284
|||||
2651 GGCTATTATGCGCAACACACATGCTATGCTTATTAACGGCTCTGTTTCC 2602
|||||
284 helLeaAsnAsnSerAlaLysIleGlyValAlaIleAlaIleGlnSerGly 300
|||||
2601 TCATTAAACAACAGCGCTAAATAATAGTGTGACTATCGCAACGATCCGGA 2552
|||||
301 GlySerLeuSerIleLeuAlaGlyGlyGlySerValLeuPheGlnAsnAs 317
|||||
2551 GGGAGTCTCTCTATCTCTGCGAGTGAAAGATCTGTTCTTCAGAAATAA 2502
|||||
317 nSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeuG 334
|||||
2501 CTCCCAACGACACCTCCGACCAAGCTCTAGTAAGAAAGCCATCTACTTAG 2452
|||||
334 IuLysAspAlaIleLeuSerSerLeuGlnAlaArgAsnGlyAspIleLeu 350
|||||
2451 AGAAGAATGCGATCTTCTCTCTTCTTAAGCTCGCAACGAGATATCTT 2402
|||||
351 PhePheAspProIleValGlnGlySerSerSerLysGlnSerProLeuPr 367
|||||
2401 TTCTTTGATCTTATTGTACAAGAAAGTACGCAAGAAATGCGCTTCC 2352
|||||
367 oSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAlaSerP 384
|||||
2351 CTCCTCTTTCAGACCAAGCTGACTTCTCCACCCCAAGCCACCTCATCTC 2302
|||||
384 rOleuValIleGlnThrSerIleAsnArgSerValIlePheSerSerGlu 400
|||||
2301 CTTTAGTATTTCAGACCAAGCTGCAACCGTTCAAGTATTTTCTCGAGCGAA 2252
|||||
401 ArgLeuSerGlnGlyGlnLysThrProAspAsnLeuThrSerGlnLeuG 417
|||||
2251 CGTCTTTGTGAAGAAAGAAACTCTGATTAACCTCACTTCCCAACTACA 2202
|||||
417 nGlnProIleGluLeuLysSerGlyArgLeuValLeuLysAspArgAlay 434
|||||
2201 GCAGCCTATCGAATCGAATCCGACGACCTTAGTTTAAAGATCCCGCTG 2152
|||||
```



```
373 ..GATTTCCTCTCTACGACAACG...GTTACTACAGCTCAAGGACCT 416
130 uArgnaSnnglySerMetSerPheCysArgAsnHisAlaGluGlySerG 147
417 TTCCCTACAGCAGACCGCTAAATTTA.....GAAATATTTC 451
147 lyGlyAlaIleSerAlaSpAlaPheSerLeuGln..... 158
452 GTAAACTGTAGTGTGCGGAATTTTTCATCTGCAGATGCTGGAGCTATCA 501
159 .....HisnTyrLeuPhe.....ThrAla.PheGluG 168
502 AAGGAGCGCTTTCTCTTAACGCGACCTTCGAGATGCTCTTTTATGA 551
168 luAsnSerSerLysGlyAsnGlyAlaIleGlnAlaGlnPheSer 184
552 ACAACTCTTCATCAACAAGGAGAGCAATTCCTACTACAGCAGCCGCT 601
185 ..LeuSerArgAsnValSerProIleSerPheAlaArgAsnArgAlaAs 200
602 CGCATAGCAAAATMACAGGTTATGTAGATTCTATCTACATAGCCGTC 651
200 pleuAsnGlyAlaIleCysCysSerAsnLeuIleCysSerGlyAsnV 217
652 TACGTCAGAGAGCGCTATCGATGATCAAGCCACGCTCATACTATGCA 701
217 AlaSnProLeuPhePheThrGlyAsnSerAla...ThrAsnGlyAla 232
702 ACAATTTCTATATTGTGAAGCAATGACGCAAAACTACGCGGTCG 751
233 lleCysCysIleSerAspLeuAsnThrSerGluLysGlySer.....le 247
752 ATCTGC.....AACACAAAGCGAGTGTCTCTCTGATC 786
247 userLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysG 264
787 GATTAACTCTACAAATTAACACTCTGATCTTCTTCAACGTAAGCAGAA 836
264 luLysGlyAlaIleTyrAlaLysHisMetValLeuArgTyrAsnGly 280
837 CAAGCGGTGGCCCATCCATCTCAAAAGCTACCCCTTCTCTCGACAGC 886
281 ProValSerPheIleAsnAsnSerAlaLysIle.....ClycI 293
887 TTTCAGAGTTTCTACGAATATGTCTCATCAGCAACTCTTAAGGGGG 936
293 yAlaIleAlaIleGlnSerGlyGlySerLeuSerIleLeuAlaGluG 310
937 TCGTATCAGATCGATGCTCTCAGAGAGCTCATCTTTCTCGACAGACAG 986
310 lySerValLeuPheGlnAsnAsnSerGlnArgThr.....SerAspGln 324
987 GAAACTTACCTTTGTAAGAAATACCTTACAAACCGAAGTACCCGAT 1036
325 GlyLeuValArgAsnAlaIleTyrLeuGluLysAspAlaIleLeuSer 341
1037 ACTCTTAACGTAATGCGATCAACTAGGAGCTAAGGGAATTCACGGA 1086
341 rLeuGlu...AlaArgAsnGlyAspIleLeuPhePheAspProIleValG 357
1087 ATTACCGGCTGCTAAAAATCATACATTTTCTTCTATCATCCCATCAGT 1136
357 lngLysSerSerLysGlySerProLeuProSerSerLeuGlnAlaSer 373
1137 CAGAAGCAACTCATCAGACGCTATTUACAGATAATTAACGCTCTCGCGGA 1186
374 ValThrSerProThrProAlaThrAlaSerProLeuValIleGlnThr 390
1187 GCTCTCAATCCATATCAAGGAAGC..... 1210
390 rAlaAsnArgSerValIlePheSerSerGluArgLeuSerGluGluG 406
1211 .....ATTCTATTTTGTGAGAAACCTTAACAGCAGATGAAC 1247
407 ..LysThrProAspAsnLeuThrSerGlnLeuGlnGlnProIleGluLeu 422
1248 TTAAAGTTGCTGACAAATTTAAATCTTCATTCACGACGAGTCCCTTA 1297
423 LysSerGlyArgLeuValLeuLysAspArgAlaValLeuSerAlaPro 439
1298 TCCGAGGAAGTTTCTCTCAAAAGGAGTCACTTTAGAGAGCAGAG 1347
439 rLeuSerGlnAspProGlnAlaLeuLeuIleMetGluAlaGlyThrSerL 456
1348 CTTCCTCGACAGAGCGCGTCTCTCTCGCATGATTCAGAGACGACAT 1397
456 euLysThrSerSer...AspLeuLysLeuAlaThrLeuSerIleProLeu 471
1398 TATCAACTACAGCTGGAGTATTACAAATCAGCAACCTAGATCAATGT 1447
472 HisSerLeuAspThrGluLysSerValThrIleHisAlaProAsnLeu 488
1448 GACTCTTAAGCTTTAAGCAGCGCTCAGCTTACAGCAAAAGTGCTTC 1497
488 rIleGlnLysIlePheLeuSerAsn.....serGlyAspG 500
1498 A...AAPAAGTGAATCTATGTGCGAAGCTCAACTGATGTATTTGAAG 1544
500 luAsnPheGlyGluAsnValGluLeuLeuSerLysGluGln..... 513
1545 GGAACATTATGAAGT...CATATGTTACAGCAAGACAGCTCTTCTCT 1591
514 .....AsnAsnIleProLeu 519
1592 CTATTAAATACACGCTTGATGCTGATGATACAGTTGACATGAG 1641
519 uThrLeuSerLysGlnSerHisLeuHisLeuProAspGlyAsnLys 536
1642 CAGCCTT.....ATCCCTGTCTCTGAGATGCTA 1673
536 erSerHisPheGlyTyrGlnGlyAspThrPheSerTrpLysAspSer 552
1674 ATTCAGAAATACGATTCACAGACAAATGTAATGTAGTACCTAGGAT 1723
553 AspGluGlyHisSer.....LeuIleAlaAsnThrTrpProLysAsn 567
1724 ACAGCTACAAATACAAAGAGCCACGCACTTGACCAAAACGAGAT 1773
567 rValProHisProGluArgGlnSerThrLeuValAlaAsnThrLeuTrp 584
1774 TGTCCAGAGCCCGAAGAAATCTGCGTTAGTATGCAATACCTATGCG 1823
584 snThrTyrSerAspMetGlnAlaValGlnSerMetIleAsnThrIleAla 600
1824 GAGTCTTACTGACATCTGCTCTGCAACAGCTTGTAGAGATCGGCGCA 1873
601 HisGlyGlyAlaTyrLeuPheGlyThrTrpGlySerAlaValSerAsn 617
1874 ACTGATATGACACACAACAGTTTCTGCTTCTTCATACAGAAC... 1921
617 uPheTyrAlaHisAspSerSerGlyLysProIleAspAsnTrpHisLys 634
1922 ....TTCCGTCATTAAGACTGAGATGAATAATCGCAAGGCTTCGCTCA 1967
634 rGlySerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspHis 650
1968 CCTGTGAGGCTACGTCAGCGTGAAGGTCTCACACACTCTTAAGAAGCAG 2017
651 SerPheCysLeuAlaAlaGlyGlnLeuGluGlyLysSerSerAspSer 667
2018 CTATTACCTTTGCGTCTGCGCATCTCTTCTGCTAGAGCAAAAGATTGTT 2067
667 eIleThrSerThrGluThr...ThrSerTrpIleAlaThrVal..... 680
2068 TATATCGCTCACACAACAGCTAGAACSTAGACSTGAGAACCTTATCTTCA 2117
```

895 CGTATCTATGCTAACGGAATGTCACATTTTCTAATATACAGCAGCTGA 944
245
945 CTTTCCAAACATATACAGCATCTCCACAACATCTCTACTGACAGCTACA 994
245
995 CCTCCACCTACACACAGCAGTACCTCTTGTAGATATGACAGCCC 1044
246 SerLeuS 248
1045 CATCTTCTGACTCCTCCAGCTACCCCCCACAACAGGTGTTACCCCTGA 1094
248 erLeuAlaCysasnIngluThrLeuPheAlaSerAsnSerAlaLeuSgl 264
1095 CTATATCTCGAAGAAACAGCCTTACATTCCTAGAAACATTCCTCCCAA 1144
265 LysGslGlyAlaIleTyrAlaLysHisMetValLeuArgTyrAsnGlyPr 281
1145 CAACGACGAGCCCTCTATGCCAAAAGATCTCTATAGATTTCTAATAATC 1194
281 ovalSerPheIleAsnAsnSerAlaLysIleGlyGlyAlaIleAlaIleG 298
1195 TACATATATTTCTTGAAATACAGCTGGAAGAGCGCTATGTCTATTC 1244
298 InSerGlyIleSerLeuSerIleLeuAlaGlyGlySerValLeuPhe 314
1245 CCGATCTCGCGAGAGCTCTCTATGCCAAATCAGGTGATTCCTCTTT 1294
315 GluAsnAsnSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaI 331
1295 AACAGAACCTCAGCATCTACTGTGGACACT...ACTGCCAATACTAT 1341
331 ePyrLeuGluLysAspAlaIleLeuSerSerLeuLualaGAsnGly 347
1342 TCACCTTCGAAAAGATGCCAAGTTGCCACTCTAGGAATACGCAAGGCT 1391
348 AspIleLeuPheAspProIleValGlnGlnSerSerSerLeuSgl 363
1392 ATACCCCTATACCTTATATCCGATTTACATCGATGATTTATCTGCTGA 1441
364 SerProLeuProSerSerLeuGlnAlaSerValThrSerProThrProAl 380
1442 TCC..... 1444
380 aThrAlaSerProLeuValIleGlnThrSerAlaAsnArg..... 393
1445 GCAGCCGCTACTGTGTCGTCATCCAAAGCCAGTGCAGATGTCGCT 1493
394 SerValIlePheSerSerGluArgLeu...SerGluGlu 406
1494 ATTCAAGGAGCTATGTCTTTTCAGAGAAACCTCAGCTACGCCAAGCA 1543
407 LysThrProAspAsnLeuThrSerGlnLeuGlnProIleGlnLeuL 423
1544 GCAACCCCTGCAAAATGCTACTACATTAACCAAAAGCTAGAACTTGA 1593
423 sSerGlyArgLeuValIleLeuLysAspArgAlaValLeuSerAlaProSerL 440
1594 AGGCGGTACTCTGCTTTAAGAAAGGTGCTACTTAATGTTCATTA 1643
440 euSerGlnAspProGlnAlaLeuLeuIleMetGlnAlaGlyThrSerLeu 456
1644 TCACGCAAAATGAAAAGTCCGTCATCATCATGATGACAGGACCAATTA 1693
457 LysThrSerSerAsp.....LeuLysLeuAlaTh 466
1694 CCACTACAAATGAGAGCTAATACTGACCGCTGCTACCTTAACAA 1743
466 rLeuSerIleProLeuHisSerLeuAspThrGluLysSerValThrIleH 483
1744 GCTTGATCATCTGGATTTCTTGATGGCACTAAAGCGGCTGCTTA 1793

483 lAlaProAsn.....LeuSerIleGlnLysIlePheLeuSer 495
1794 ATGTGACAGATACCAATGGAGCTCTCACTATATCCGAGACTTTAGACTT 1843
496 AsnSerGlyAspGluAsnPheTyrGluAsnValGluLeuLeuSerLysG 512
1844 GTGAAAACCTCAAGATTGCTGTGACAAACCCAGGAGATGTTAATAAGA 1893
512 uGlnAsnHisIleProLeuLeuThrLeuSerLysGluInSerHisLeuH 529
1894 TTTCAGCAGACTCCGATTTTGAAGCTCAAGCAGACTTCAATACTGTA 1943
529 lAlaProAsp.....GlyAsnLeuSerSerHisPhe 539
1944 CCAGTACGAGACTTCAGTCTCGCAAAACGGCTATCAGCAATTCGCCAT 1993
540 GlyTyrGlnGlyAspTyrThrPheSerThrLysAspSerArgGluGlyH 556
1994 GGTATACAGGAGACTTGGAGTTTACATACAGACAGCAAC...CA 2037
556 sSerLeuIleAlaAsnThrProLysAsnTyrValProHisProGlu 573
2038 TACGGTACAGGAATTTGAAAAACCGGTTATCTCTCTATCCGAGAC 2087
573 rGlnSerThrLeuValAlaAsnThrLeuThrPAsnThrTyrSerAspMet 589
2088 CTCTTGCTGCCCTCATCTTCATATAGCCATAGCGCAACCATCATGATTTA 2137
590 GlnAlaValGlnSerMetIleAsnThrIleAlaHisGly..... 602
2138 CGAGCTGTGAAGTCAAGC...TCAGCAGCTGATGGCGAAGATGTC 2181
603 GlyAlaTyrLeuPheGlyThrTyrGlySerAlaValSerAsnLeuPhe 619
2182 TGGAGAGCACTGAGACATCACA.....GGAATTCAAATTTCTTCC 2222
619 yAla...HisAspSerSerGlyLysProIleAspAsnThrHisHisArg 634
2223 ATCGAATCAATACCGGTATGACAGC...ACGTACCGCATATG 2263
635 SerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAsnHis 651
2264 GGTGGAGGCTACTTCATCAATACCTACACAGCATCACTCCAGATGTC 2313
651 rPheCysLeuAlaAlaGlnLeuLeuGlyLysSerSerAspSerPhe 668
2314 GTTAGTCTAGGTTTGGACAGACTGTTACAAAATCAAGATTACCTCG 2363
668 lThrSerThrGlnThrThrSerTyrIleAlaThrValGlnAlaGlnLeu 684
2364 TAGTCACGATCATCTTAACGTTTATTCGTCACAGTATACCTTAACATC 2413
685 AlathrSerLeuMetLysIleSer.....AlaG 694
2414 ACCAAGTCTCGTTTGATCATCGAATCTTCTCAGAGAGCACTTCG 2463
694 nAlaCysTyrAsnGlnSerIleHisGluLeuLysThrLysTyrArgSerP 711
2464 AGTTACCTATACCGCTACCAATGAGAAAGTAAGCTTCATATACAAAT 2513
711 heSerLysGluGlyPheGlySerThrHisSerValAlaValSerGlyL 727
2514 TCCCTAA...GGCGCGTCTCTTGAGTACCAATTCGTGTTAGAGAA 2560
728 ValCysAlaSerIleProIleValSerAsnGlySerGlyLeu...Phe 743
2561 CTCGAAGGAACTTCCATCACTCTCTTCCATCTTAACCTCA 2610
743 rSerPheSerIlePheSerLysLeuGlnGlyPheSerGlyThrGlnAsp 760
2611 GCAATCATCTCCCTTTGTAAGCTGAAGTTCCTTACGAGCTCATGGG 2660


```
578 CCGCCGAACAAGGGGACCGATCTAAGCTTTCGTCGACTCTGATTTCT 627
187 ArgAsnValSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnI 203
628 GAAATGCGAGGAAATCTTGAGCTTGGAACAACAGTGCACACATCAAG 677
203 yGIyAlaIleCysCys...SerAsnLeuIleCysSerGIyAsnValAsp 219
678 AGCCCGATCTCTGCTGAAGGACCTTGATCTTCATTAACCAAAAT 726
219 roleuPhePheThrGIy...AsnSerAlaThrAsnGIyAlaIleCys 234
727 ..ATCTTTTCGATGCTGCACAAAGCACTAACAAATGCGGACTATTCAT 774
235 CysIleSerAspLeuAsnThrSerGIyLysGIySerLeuSerLeuAla 251
775 TCT...AACAAAGCAGGGGCGAACCCAGACCTATCTTACCTTTCAGG 821
251 sAsnGIyLysThrLeuPheAlaSerAsnSerAlaLysGIyLysGIyAla 268
822 AATGAGAGCCGTCATTTCTGATATACACACAGCAATATGAGAGTG 871
268 IaIleTyAlaLysHisMeValLeuArgTyr...AsnGIyProValSer 283
872 CGATTTAATACCAAAAATTCGTGTATCCTCAGACAGCAGCAGCTTCA 921
284 PheIleAsnAsnSerAla.....LysIleGIyAlaIleAla 296
922 TTTTCTAACACAAAGCTGCGAATCCTACTCCTAAAGAGGGCATATTC 971
296 aIleGIySerGIyLysSerIleLeuAlaGIyLysGIySerVal 313
972 GATTCAGATTCGTGAGAGATTAGCATTTCTGACATCTCGGCATATCA 1021
313 euPheGIyAsnAsnSerGIyAlaArgTyr...SerAspGIyLysLeuAla 327
1022 TTTTCCAGGCGCAATCTACGACACTACAGCAAGTCGCGAGTGAGAC 1071
328 ArgAsnAlaIleTyLeuGIyLysAspAlaIleLeuSerSerLeuAla 344
1072 ACAAATGCTATGATCTTGATGATGAAATTTTAAATCTCGGAGC 1121
344 aArgAsnGIyAsp...IleLeuPhePheAspProIleValGIyLys 360
1122 GACCTGGGGAAATAAGTTATTTCTATGATCTCTATC..... 1158
360 eSerLysGIySerProLeuProSerSerLeuAlaSerValThrSer 376
1158 ..... 1158
377 ProThrProAlaThrAlaSerProLeuValIleGIyThrSerAlaAsn 393
1159 .....ACGAGCTCAGAGACTGATAGCTCTCTTTGAAATTA 1196
393 gSer.....ValIlePheSers 399
1197 AGCTACGCGAGATCTGGAATACCTATGAAGCTACATCTTTCTCTCG 1246
399 eArgIaArgLeuSerGIyLysLys...LysThrProAspAsnLeuThrSer 414
1247 GAGACAACACTCTCAGAGACGAACTTAAGAAACCTGACATCTGAAGCT 1296
415 GIyLeuGIyLysProIleGIyLysSerGIyArgLeuValLeuLysAs 431
1297 ACATTTACACAGGCTGTAGAGCTTCTGAGGTGCTTATTAATTAAGA 1346
431 pArgAlaValLeuSerAlaProSerLeuSerGIyAspProGIyAlaLeu 448
1347 TGAATGACTGCTAGTTGCAATATACAGAGCTCGAGGATCGAAG 1396
448 euIleMeGIyAlaGIyThrSerLeuLysThrSerSerAsp...LeuLys 463
1397 TCGTTATGATGAGAGGAGCTACTTTTTCAGGCAAGCGCTGAGGCGTCA 1446
464 LeuAlaThrLeuSerIleProLeuHisSerLeuAsp...ThrGIyLys 479
1447 CTCATGCGCTAGCCATTATATAGATTCCTTAATGAGCAAAATAAGC 1496
479 rVal.....ThrIleHisAlaLeuAsnLeuSerIleGIyLysIleP 493
1497 TATCATTAAGGCGACCGCCAGCAAGTACAGATGCTGCTTATACGGCC 1546
493 heLeuSerAsnSerGIyAspGIyAsnPheTyGIyLysAsnValGIyLeu 509
1547 TCATGCTTGTAGATGCTGCGGGAACTATTATGAGATCATATCTC... 1593
510 SerLysGIyLysAsnAsnIleProLeuLeuThrLeuSerLysGIy... 524
1594 ...AGTCAACACAGAGTCTTCTTAAATGAGCTTTCGCAACAAGAAC 1640
525 ..GIySerHisLeuHisLeuProAspGIy.....AsnLeuSerHis 538
1641 GATGACTACTACAGATATCCCGATACCCCATTTCTAAATACTACGAAT 1690
538 IaPheGIyTyGIyLysAspTrpThrPheSerTrpLysAspSerAspGIy 554
1691 ACATAGGATATCAGAGAACTGCAATATATGTTGGGTGACAGATGCAACT 1740
555 GIyHisSerLeuIleAla.....AsnTrpThrProLysAsnTyValPr 569
1741 GCAAAAACAAAAAATGCTACTTACTTGAAGCTTAACAGATACAAAGCC 1790
569 oHisProGIyArgIleSerThrLeuValAlaAsnThrLeuTrpAsnThr 586
1791 GATTCAGAGACTCAGAGCACTTTGGTTCCTTAATAGCTGTGGGTTCTT 1840
586 TyrSerAspMeGIyAlaValGIySerMetIleAsnThrIleAlaHisGIy 602
1841 TTGTGATGATGCCGCTCCATTCAGACCTCATGACCGGAGCAAGTTGC 1890
603 GIyAlaTyLeuPheGIyThrTrpGIySerAlaValSerAsnLeuPhe 619
1891 TTAATCTTGTCACAAATTTGTGGTATCAGAAATCGCGGAC...TTT 1937
619 rAlaHisAspSerSerGIyLysProIleAspAsnTrpHisIleArgSer 636
1938 GCATGAAGATCAAGAAAGAACCA...CGTATGATATGCTATTTCTAGC 1984
636 euGIyTyLeuPheGIyLysSerThrHisSerLeuAspHisSerPhe 652
1985 CGGCTTATGATTAAGAGAGATTCCTCACGCTTGAAATTTCTT 2034
653 CysLeuAlaIleGIyLysLeuLeuGIyLysSerSerAspSerPheIle 669
2035 AATTTGCTTTTGTGACCTTTTGGCTACGACAAGACATCTGTGGC 2084
669 rSerThrGIyThrThrSerTrpIleAlaThrValGIyLysIle 683
2085 TAAGAACCATACCCATGTATATGACAGGGCAATGATACCGACACCTCG 2134
684 ..LeuAlaThrSerLeuMetLysIle..... 691
2135 GAGACTTAAGACCCCTCGCTAGATTTGTGAGAAATTCGACTCCCTA 2184
692 .....SerAlaGIyAlaCysTyArgAsnGIyLysIleHisGIyLeu 704
704 LysThrLysTyArgSerPheSerLysGIyLysPheGIyLysTrpHis 721
2235 GACCAACAAGTACACTGGCTATTTCTCTGTTAAG...GGAAGCTGGGAA 2281
721 eValAlaValSerGIyLysValCysAlaSerIleProIleValSerAsn 737
2282 ATGATGCTCTGCTATGAATGAGAGAGATATCCGTTAGTTCTTCA 2331
```

```
487 GGGCCCATCATCGCTAAATACACACCCCTCAGCAAAAGCTGCATTAAAT 536
124 yAsnGlyGluLeuSerLeuArgAsnAsnGlySerMetSerPheCysArg 140
537 GTCGAGGGGATCTTACATTTCATACAAATGCAATATTTTATTATAACAA 586
141 AsnHisAlaGluGlySerGlyAlaIleSerAlaAspAlaPheSerLe 157
587 GATTACTGTAGAAATAGCGGAGCCATTCTCCAGAAATCTTCTTCTT 636
157 uClnHisAsnTyrLeuPheThrAlaPheGluGluAsnSerSoriGly 173
637 GAAAAACACGACGCGATCATTTCTTTGAAAGCCAAATTAATCCACC 686
174 ..... AsnGlyAlaIleGlnAlaGln... ThrPheSerLeuSer 186
687 CAGCAAAAAGGTGGCCCTATTTGCTACTGTAGTACTAGATATATACA 736
187 ArgAsnValSerProIleSerPheAlaArgAsnArgAlaAspLeuAsn 203
737 AATATATAGGCTCTACCTCTCTCCGAAACAATATGCTGAAGCTGCA 786
203 yGlyAlaIle... CysCysSerAsnLeuIleCysSerGlyAsnValAsp 219
787 TCGAGCTTAAATAGCACAGCAAACTCTACAAATATACAGCAATACCT 835
219 rOlLeuPhePheThrGlyAsnSerAlaThr..... AsnGlyGly 231
836 ..CTGTATTTCTGAAATATAGTGTGACAGCAGCCAGCAAAATCCAGA 883
232 AlaIleCysCysIleSerAspLeuAsnThrSerGluLysGlySerLeu 248
884 GCTCT..... TCTGGAGATGCCGATGTAC 909
248 rLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysGlu 265
910 CATATCTGGAGATCAGACTTACTTCTCAGAAACCACTGACTAGCTA 959
265 ySGlyGlyAlaIleTyrAlaLysHisMetValLeuArgTyrAsnGly... 280
960 ATGGCGAGCCATTATCTAGAAAGCTTACACTGCTCCGGGGGGGG 1009
281 ..... ProValSerPheIleAsnAsn..... SerAl 289
1010 GGGGGGAATCCC..... TTTCTAACATATAGTCCAAAGTACCACATCC 1053
289 aLysIleGlyAlaIleAlaIleGlnSerGlyGlySerLeuSerIleL 306
1054 AGTAATGCTGAGCCATTTCTATCTAGCTGACAGCTGAGTGTACTT 1103
306 euAlaGlyGluGlySerValLeuPheGlnAsnAsnSer... GlnArgThr 321
1104 TCAGCGAAGCAGGGGACCATTAACCTTAATGGAAATCCCATTTGCAACT 1153
322 SerAspGlnGlyLeuValArgAsnAlaIleTyrLeuGluLysAspAla 11 338
1154 ACACACACAACTACAAAAGAAATTTCTATTCATAGATCTACTGGCAA 1203
338 eLeuSerSerLeuAlaArgAsnGly... AspIleLeuPhePheAsp 354
1204 AGATCACAATTAAGTGCATATCTGGCATAATCATTTTCTTACAGATC 1253
354 rOlLeuAlaGlnGlySerSerLysGlySerProLeuProSerSerLeu 370
1254 CGATTACTGCCAATACGCTCCGATCTCTACAGTACTTTTAATCTCAAT 1303
371 GlnAlaSerValThrSerProThrProAlaThrAlaSerProLeuVal 11 387
1304 AAGCGTATGAGGTAAAT..... 1321
387 ecInThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerG 404
1322 .AGTACACATTATAGTGGGTGATTGTTTTTCTGCTCAAAAAGCTCTCTG 1370
404 LuGluGlu... LysThrProAspAsnLeuThrSerGlnLeuGlnPro 419
1371 AAGATGAAGCAAAAGCTTCAGCAACCTCCTACTTACCTCAGAGAGCT 1420
420 IleGluLeuLysSerGlyArgLeuValLeuLysAspArgAlaValLeu 436
1421 GTACTCTTAAGTGCAGAAATTTAGTACTTAACCTGCTGTACCTCTGA 1470
436 rAlaProSerLeuSerGlnAspProGlnAlaLeuLeuIleMetGluAla 453
1471 TACGAAAGCCTTTACTACTAGACGGCGGCTCCCTCTTATTATGATGGG 1520
453 lYThrSerLeuLysThrSerSer... AspLeuLysLeuAlaThrLeuSer 468
1521 GCACAACTTAAAGCAAGTACAGAGAGGTACACTTATACAGGCTTTCC 1570
469 lLeProLeuHisSerLeuAspThrGluLysSerValThrIleHisAla 485
1571 ATTCTGTAGACTCTTATGGCGAGGATAGAAAGTTGTAATGTGCTTTC 1620
485 O..... AsnLeuSerIleGlnLysIlePheLeuSerA 496
1621 TGCAAGCAAGTAAAAATAGCCCTTAGCGTCCTCCTCTGCTGCTACGAA 1670
496 snSerGlyAspGluAsnPheTyrGluAsnValGluLeuLeuSerLysGlu 512
1671 ACCAAGG..... AATGCTTATGAAATCAGAC... TAGGAAAAACT 1711
513 GlnAsn..... AsnIleProLeuThrLeuSerLysGlu 525
1712 CAAAGCTTTTATTGTCACCTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 1761
525 nSerHisLeuHisLeuProAspGlyAsnLeuSerSerHisPheGlyTyr 542
1762 AGATGTCACAGCGGCTCTACAGTACCAATCTCAGCAATGCTGATC 1811
542 lngLysPheThrPheSerThrLysAsp..... SerAspGlu 554
1812 AAGTACTTGGGAATGAGTGGTGTATGATACCGCAAGCAGCCCAAG 1861
555 GlyHisSerLeuIleAlaAsnThrProLysAsnTyrValProHisP 571
1862 ACTAAGACAGGACATTAGCTTGAGCAATACAGCTTCCGGAATCC 1911
571 oGluArgGlnSerThrLeuValAlaAsnThrLeuThrAsnThrTyrSer 588
1912 TGACGCTCAAGACCTTATAGTCTCAATAGCCTTGAGGATCTTTTCA 1961
588 spMetGlnAlaValGlnSerMetIleAsnThrIleAlaHisGlyAla 604
1962 ACATCCAAAGCATTCAGGTGTCTATAGAGAAAGTGTGACTCTTGT 2011
605 TyrLeuPheGlyThrPGLYSerAlaValSerAsnLeuPheTyrAla 621
2012 TCAATGAGGCTCTGAGCGCGGAGTCCCAAT... TTCTTAGATAA 2058
621 sAspSerSerGlyLysProIleAspAsnThrPHisHisArgSerLeu 638
2059 AGATAAGAAAGGGAATAA... CCAATATACGTCATAAATCTGCTGAT 2105
638 yrLeuPheGlyIleSerThrHisSerLeuAspAspHisSerPheCysLeu 654
2106 ATGCTATCGAGGTGAGCGCAAACTGTCTGAAACTTAATATAGCTTT 2155
655 AlaAlaGlyGlnLeuLeuGlyLysSerSerAspSerPheIleThrSer 671
2156 GCCTTTGGCAACTCTTGTGAGGATAAAGATTTCTTACGCTTAATA 2205
671 rGluThrThrSerTyrIleAlaThrValGlnAlaGlnLeuAlaThrSer 688
2206 TCATACTGATACATGACAGAGAGCTTCTATATCCACACATTACAGAT 2255
```


1005 GAAATATATGTCATCAGCAACTCCTAAGGGGGTCTCTATCAGATCGAT 1054
299 SerGlySerLeuSerIleLeuAlaGlyGlySerValLeuPheG1 315
1055 GCGTCAGAGAGCTCAGCTCTTCTCGAGACAGACAGAAACATTACCTTTGT 1104
315 nanaSerSerGlnArgThr.....SerAspGlnGlyLeuValArgAsn 330
1105 AACAAATACCTTACACACACCGAATGCCGATATCTCTAAACGTATG 1154
330 JalIerLeuGlyLysAspAlaIleLeuSerSerLeuGlu...AlaArg 345
1155 CGATCAACATAGGAGACTAACCGGAAATTCACCGGAATTCACCGCTCCATAA 1204
346 AsnGlyAspIleLeuPhePheAspProIleValGlnGlnUserSerLeu 362
1205 AATCATCAATTTCTTCTATGATCCATCATCATCAAGACGACCTCATC 1254
362 sGlySerProLeuProSerSerLeuGlnAlaSerValThrSerProThr 379
1255 AGACGTATTGAGATTAATAAGGCTCTCGCGAGCTCTCAATCATATC 1304
379 roaIarThrAlaSerProLeuValIleGlnThrSerAlaAsnArgSerVal 395
1305 AAGGACG.....ATP 1315
396 IlePheSerSerGlyArgLeuSerGlnGlu...LysThrProAsp 411
1316 CTATTTCTCGAAGAACCTTAACAGACATCAACTTAACCTTCTGCAAA 1365
411 nLeuThrSerGlnLeuGlnProIleGlyLeuLysSerGlyArgLeu 428
1366 TTTAAATCTTCAATTCACGCGACCACTCTCCGAGGAAAGTAT 1415
428 aLeuLysAspArgAlaValIleuSerAlaProSerLeuSerGlnAspPro 444
1416 TCGTCAAAAAGGACGACTTTAGAGACACGACCTCTCTCAAGAGCC 1465
445 GlnAlaLeuLeuIleMetGlyAlaGlyThrSerLeuLysThrSerSer 460
1466 GCTTCTCTCCGCGCATGGATTGACAGAACGACATTATCACTACAGCTGC 1515
461 AspLeuLysLeuAlaThrLeuSerIleProLeuHisSerLeuAspThrG 477
1516 GAGTATTTCATTCACGAACCTTAGATCAATGTGACTCTTAGTCTTA 1565
477 LulysSerValThrIleHisAlaProAsnLeuSerIleGlnLysIlePhe 493
1566 AGCAGCCCGCTCACCCCTAACGAAAGCGTCTCA...AATAAGCTGATC 1612
494 LeuSerAsn.....SerGlyAspGlnAsnPheTrpGluAs 505
1613 CTATCTGGGAAGCTCAACCTGATTCATTTGAGAGACATTATGAAAG 1662
505 nValGlyLeuLeuSerLysGlnGln..... 513
1663 T...CATTTGTCAGCCATGACAGCTCTCTCTATTAAATACAGG 1709
514AsnAsnIleProLeuLeuThrLeuSerLysGlu 524
1710 TTGATCTGATGTTGATTAACCTTGCATTCAGACGACCTT..... 1750
525 GlnSerHisLeuHisLeuProAspGlyAsnLeuSerHisPheGlyTy 541
1751ATCCCTGTTCTGCTGAGGATCCTAATTCAGATAGCGATT 1791
541 rGlnGlyAspTrpThrPheSerTrpLysAspSerAspGlnGlyHisSer 557
1792 CCAAGGACAATGGAATGTTATGACATACGATACAGCTACAATAACAA 1841
558LeuIleAlaAsnTrpThrProLysAsnTrpValProHisProGlu 572
.....

1842 AAGAGCCCGACGGCACTTGACCAAAACAGGATTTGTTCCACCCCGAA 1891
573 ArgGlnSerThrLeuValAlaAsnThrLeuTrpAsnThrTySerAspMe 589
1892 AGAAATCTCGGTTAGTATGCATATACCTTAGGGAGATCTTACTGAT 1941
589 rGlnAlaValGlnSerMetLeuAsnThrIleAlaHisGlyAlaTyL 606
1942 TCGTCTCTGACACAGCTTAGAGATCGGCGCAACGATGAGACCA 1991
606 ePheGlyThrTrpGlySerAlaValSerAsnLeuPheTyAlaHisAsp 622
1992 AACAGGTTTCTGGGTTTCTCCATGACGAGC.....TTCCTGCAATAG 2035
623 SerSerGlyAspProIleAspAsnTrpHisHisArgSerLeuGlyTyL 639
2036 ACTGAGATGAAATTCGCAAAAGCTTCGTCATACCTCGAGGCTAGT 2085
639 uPheGlyIleSerThrHisSerLeuAspAspHisSerPheGlyLeuAla 656
2086 CATCGGTGAAGTCTCACACTCCTAAAGACGACCTTTTACCTTTCGCT 2135
656 IaGlyGlnLeuLeuGlyLysSerSerAspSerPheIleThrSerThrGlu 672
2136 TCTCCCATCTCTTGTGACAGACAAAGATTTGTTATGCTCACACAAAC 2185
673 ThrThrSerTyIleAlaThrVal..... 680
2186 TCTAGAACCTAGCGTGAACCTTTTCTCAAGACCTCTCATACCCTACA 2235
681GlnAlaGlnLeuAlaThrSerLeu 689
2236 ACCCAAAACTATTGAGATTAGAGAGCAAAAGTTTCTGTAACACTA 2285
689 eLys.....IleSerAlaGlnAlaLys 696
2286 TAGAAAATTCCTCAAGGAAATTCCTACGCTTGATGCCAAGTTTCG 2335
697 TyArgGlnSerIleHisGlyLeuLysThrLysTyArgSerPheSer 713
2336 TTCAGCCATTGACACACCGTATGGAACGACATATACCTGATTCG...CC 2382
713 sGlyGlyPheGlySerTrpHisSerValAlaValSerGlyGlyLys 730
2383 AGAATCCGAAAGTTCTTGACCAACGAGTGTATGCTGATGCTGCGCC 2432
730 IAserIlePro...IleValSerAsnGlySerGlyLeuPheSerPhe 745
2433 TAGACCTTCTTGTGTTCTTCCAAACCATCTCTTTTCAACACCTTC 2482
746 SerIlePheSerLysLeuGlnGlyPheSerGlyThrGlnAspGlyPheG1 762
2483 ATTCACAGATGAAAGTGAATGTTATGATACACAAATATAGCTTCT 2532
762 uGlySerSerGlyGlyLeuArgSerPheSerAlaSerPheArgAsn 779
2533 CGAAAGCTCTAGTATGCGCGTGTATGATGAGAGGCTGCTTAC 2582
779 IAserLeuProIleGlyThrPheGlyLysLysSerGlnLysThrArg 795
2583 TCTGATTTCTGTTGGTGGGAAATTCGTCAGGGGATATGCGAGATTCC 2632
796 ThrTyTyTyPheLeuGlyAlaTyIleGlnAspLysArgAspVal 812
2633 TACACCTATGATCTCTCAGATTTCTTGTTCGATGCTATGTAACAA 2682
812 IGlusSerGlyProValValLeuLeuLysAsnAlaValSerTrpAspAla 829
2683 TCCCAATCTACAGGACTCTTGTGATGAGCCACAGACTCTTGGAATTC 2732
829 rGlnAlaAsnLeuAspSerArgAlaTyLrmePheArgLeuThrAsnGln 845
2733 GCGTGCCAACTTTTCAAGACAGCATTTTACTGAGGCGTACGACAC... 2779
.....

```

1111 .....1111 111 .....111 111
1260 CCGGAACGTGACGGACTTCTGCAAGTCAAGTAACGCTATCTATTTC 1309
352 PheAspProIleValGlnGlnSerSerSerIleGluSerProLeuProSe 368
1310 TATGATCCCAT..... 1321
368 rSerLeuGlnIleSerValThrSerProThrProAlaThrAlaSerProL 385
1322 .....ACCACCAACGATACGGGAGCAAGCAAGCAATCACTTAC 1355
385 euValIleGlnThrSerAlaAsnArg.....SerValIlePhe 397
1356 GTATCAATGACGCTGACGTCAAAATCAAAAGCTTCGAGATCTATGATATTT 1405
398 SerSerGluAlaGluSerSerGlnGlnGluLysThrProAspAsnLeuThrSe 414
1406 TCTGGAGAGAGATTGTCACACAGCAAGCTTACCTGAAATCTTACATTC 1455
414 rGlnLeuGlnGlnProIleGlnLeuLysSerGlyArgLeuValLeuLysA 431
1456 GAGGATCAACACGCTGTCACTTATGATAGAGGGAGCTTACTTAAAC 1505
431 sPArgAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeu 447
1506 AUGGAGTCACTTGATGCACACAGAGATTCTCGACAGACCAAGATCCACG 1555
448 LeuIleMetGluAlaGlyThrSerLeuLysThrSerSer...AspLeuLys 463
1556 CTTCCTTTGGATCTGGGCACTCAATTAAGCTTCTACAGAAATATTTCT 1605
463 sLeuAlaThrLeuSerIleProLeuHisSerLeuAspThrGluLysSerV 480
1606 CATCAAAATTAATCTATTAATGCCGATACCATTTACGGAAGAAATCCCTA 1655
480 aThrIleHisAla.....ProAsnLeuSerIleGlnLysIle 492
1656 TCAATATTGTAGCTTCACACAGAGATTAAGACCTTACCTTAACAGCAAGC 1705
493 PheLeuSerAsnSerGlyAspGlnAsnPhetYrGluAsnValGlnLeuLe 509
1706 TTACACATTGTAATAATGACATGACCTTTCTATGAGAACCAATACCTTCCA 1755
509 userLysGlnGlnAsnAsnIleProLeuLeuThr..... 520
1756 AGACTCTCAAGACTATACCTTTGTAAAGTTATCTCCAGAGACGGAGGGA 1805
521 .....LeuSerLysGlnGlnSerHisIleuHisLeuProAspGlyAsnLeu 535
1806 CTATATATTACTCAAGATGCTCTCAGAAAGCCCTTGAAGTACCTCTCTCT 1855
536 SerSerHisPheGlyTyrGlnGlnLysPThrPheSerTrpLysAspSe 552
1856 AGACCACTATGCTATCAAGACATTCGAAATTCGCAAGTCATCCACAG 1905
552 rAspGlnGlyHisSerLeuIleAlaAsn.....TrpThrProLysAsnTr 567
1906 AACGGGAACTCAACGACGCAAGCAAAATTTAAATGGGTGGCGAGACGAT 1955
567 YrValProHisProGlnArgGlnSerThrLeuValAlaAsnThrLeuTrp 583
1956 ACCCTTCGGAATCCGACGCAAGCATCTTTAGTCCCAATAGACCTTCG 2005
584 AsnThrTrpSerAspMetGlnAlaValGlnSerMetIleAsnThrIleAl 600
2006 GGTCTCTTTGTCATCAGCTCTCTATCCAGAAATATCTGTAATATAGTAG 2055
600 aHisGlyGlnAlaTyrLeuPheGlyThrTrpLysSerAlaValSerAsnL 617
2056 CCAAAATCTTATCTCAAGAAAGGAGCTCGGGAGCTGCAATTTGCTAAT 2104
617 euphetYrAlaHisAspSerSerGlyLysProIleAspAsnThrHis 633

```

```

2105 ..TTCTACATAGAGATAAATTAATGAG.....CACGGCTATCGCCAT 2146
634 ArgSerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAspHI 650
2147 AGCGGTGCGGTATCTGTGGGAGTTGGACATCATCTCTTTTCTGATGC 2196
650 sSerPheCysLeuAlaIleGlnLeuLeuGlyLysSerSerSerP 667
2197 TACGATTAATCGGCTTTTTCGAGCTCTTCAGTAGAGATAAAGACTACG 2246
667 heIleThrSerThrGluThrThrSerTyrIleAlaThrVal..... 680
2247 TACATCCAAATATCATGAGACTAGTACTCAGGGGCTGTATTCTTGAG 2296
681 ..... 681
2297 GATACCTAGAGTTTAGAAGTCACAGGAGTTTATCTGATACCTCTTC 2346
681 nAlaGlnLeuAlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrA 698
2347 AGAAGCTTGCTGTAACCAAGTCGTCACTATGATATGAGTGTCTTACA 2396
698 sngLysSerIleHisGlnLeuLysThrLysTyrArgSerPheSerLysGlu 714
2397 GCCATAGAAATATGATATGAAACCAAAATACACGACATAT...CCAGAA 2443
715 GlyPheGlySerTrpHisSerValAlaValSerGlyGluValCysAlaSe 731
2444 GCTACGAGATCTTGGGCAATGATGTTTGTGCTTGAGTTGAGAGCAC 2493
731 rIleProIleValSerAsnGlySerGlyLeuPheSerSerPheSerIleP 748
2494 TACATACACTACCCCTAACATCTATTTTATTTGATACACTCTCCGT 2543
748 heSerLysLeuGlnGlyPheSerGlyThrGlnAspGlyPheGlnGluSer 764
2544 TTCTCAGGCTGTCAGGCACTATGCTCACAGGAAGACTTCAAGAGACA 2593
765 SerGlyGluIleArgSerPheSerAlaSerSerPheArgAsnIleSerLe 781
2594 GGAGGTGAGTTGCGTCACTTACTAGCGGAGATCTTTCAATTTAGCAGT 2643
781 uProIleGlyIleThrPheGlnLysLysSerGlnLysThrArgThrTyrT 798
2644 TCCATTTGGCGTGAAGTTTGAGAGATTTTCAGCTGAAGGGGATCTCT 2693
798 YrTyrPheLeuGlyAlaTyrIleGlnAsp...LeuLysArgAspValGlu 813
2694 ATGAACCTTACCCCTTCATGTTGCTCTGATGTCGATTCGCAAGATCCCAAG 2743
814 SerGlyProValLeuLeuLysAsnAlaValSerTrpAspAlaProMe 830
2744 AGC.....ACGGCAACATTGGCTAGTGAAGCTAGGTGAGACACCCACGG 2787
830 tAlaAsnLeuAspSerArgAlaTyrMetPheArgLeuThrAsnGlnArgA 847
2788 AAACAATCTCTCCAGACAAAGATTACACTGCGTTTAGCAACCCCTGTC 2837
847 lAlaLeuHis...ArgLeuGlnThrLeuLeuAsnValSerCysValLeuArg 862
2838 TCATTAATCCCTGGAATTCGAGGTGTTCACTCAACGAGCAATTAATGCGG 2887
863 GlyGlnSerHisSerTrpSerLeuAspLeuGlyThrTrpTyrArgPhe 878
2888 GGAATCTCTCGTAATTAATTAACATCAATCTCGGGGGTAAATACCGATT 2935
seq_name: /SID52/gcgdata/geneseq/geneseqn/NA2000.DAT.A30847
seq_documentation_block:
ID A30847 standard: DNA; 2950 BP.
XX
AC A30847;
XX

```


1024 GATATGCTGACATTCCTCTTTTACGGGAATACACTCTTACTACTGCC 1073
324 nglyLeuValArgAsnAlaIleTyrLeuGluLysSerAlaIleLeuSer 341
1074 TGGGACGAAATGAGTACTATGACTTAGAACGAGTGCAGAAAGAGACAG 1123
341 erLeuGluAlaArgAsnGly...AspIleLeuPhePheAspProIleVal 356
1124 CTTTCGCTTGTGCTGCTGTAGAGCACTACTCTATGATCCCATACT 1173
357 GluLysSerSer...LysGluSerPr 365
1174 ACAGGATCTTCACACAGTACAGATCTCTAAAGTTAATGACACTCC 1223
365 oLeuProSerLeuGluAlaSerValThrSerProThrProAlaThrA 382
1224 GGCAATCTTCGACTACAAATATACA... 1248
382 laSerProLeuValIleGlnThrSerAlaAsnArgSerValIlePheSer 398
1249GGGACATCATCTTCCAC 1266
399 SerGluArgLeuSerGluGluLysThrProAsp...AsnLeuThrSe 414
1267 CGAGAAACTTATCAGACAGACAGCCCGCAGATCTTAAATACTCTAC 1316
414 rGlnLeuGlnGlnProIleGluLeuLysSerGlyYargLeuValLeuLysA 431
1317 GAAGTACTACAGCCTGTACTCTTTCAGAGAGTACTATCTTTAAAC 1366
431 spArgAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeu 447
1367 ATGGATGACCTCGACACTCAGCATTCACCTCAACAGCAGATCTCGT 1416
448 LeuIleMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLe 464
1417 CTCGAAATGAGCTAGAGACTCTAGAACCTGCTCATACTACGACCAT 1466
464 uAlaThrLeuSerIleProLeuHisSerLeuAspThrGluLysSerValT 481
1467 AACCAATTGGTCATTACATCATCAGTCTTATAGACCGTGCAGAAAGAGCA 1516
481 hrIleHis.....AlaProAsnLeuSerIleGlnLysIlePhe 493
1517 AAATGAAACCAAAAGCTAGCTCAAAAATACTGACTTATCTGGAACCTC 1566
494 LeuSerAsnSerGlyAspGluAsnPheTyrGluAsnValGluLeuLeuSe 510
1567 ACTTTATTGGACCCGACCGCAGCTTTATGAAAATCATACTTTAAGAAA 1616
510 rLysGluGlnAsnAsnIleProLeuLeuThrLeuSer...LysGluGln 526
1617 TCCTCAGTCCAGCATCTTAGAGCTCAAAAGCTTCTGGAACSTGACAA 1666
526 erHisLeuHisLeuProAspGlyAsnLeuSerSer...HisPheGly 540
1667 GCACCGCAGTACCTCCAGATCTTATAGGTGAGAAATTCATTAACGC 1716
541 TyrGlnGlyAspTrp...ThrPheSerTrpLysAspSerAspGluGlyHis 556
1717 TATCAGGGAACCTTGGCCCAATCTTTGCG.....GGCAGCGGGCG 1757
556 sSerLeuIleAla.....AsnTrpThrProLysAsnTyrValProHisP 571
1758 TTCTACGACTCCAACTCTGACTGACTAAAGTGCCTATATCTCTATC 1807
571 rGluValArgGlnSerThrLeuValAlaAsnThrLeuTyrPAsnThrTyrSer 587
1808 CCGAGCGTATCGGCTCTTACTCTTACTCTTATGAAATGCAATTTATA 1857
588 AspMetGlnAlaValGlnSerMetIleAsnThrIleAlaHisGly.... 602
1858 GATATTACCTCTCTCCATTATTTATGAGACTGCAGAAAGCGCTTGCA 1907

603 .GlyAlaTyrLeuPheGlyThrTrpGlySerAlaValSerAsnLeuPheT 619
1908 GGGAAACCGTGCTTT.....TGGTGCTGCGATTACTACTCTTCTC. 1950
619 yAlaHisAspSerSerGlyLysProIleAspAsnTrpHisIleArgSer 635
1951CAATAGGATAGTACAAAACACGACGCGGGTTCCGCATTTAGT 1995
636 LeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAspHisSerPh 652
1996 GCGCGTTATGTCATAGAGGAACCTACATCTTGTGATTAAGATTTCT 2045
652 ecysLeuAlaIleGlyGlnLeuLeuGlyLysSerSerSerPheIleT 669
2046 TAGTCTGCTATTGTGCTGCTCTTGGAGAAGATAGACTACTTTGTAG 2095
669 hrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaGlnLeuAla 685
2096 CTAAGAATCAAGTACAGCTTACGAGAGACCTTATTACAGACACAC 2145
686 ThrSerLeuMetLysIle..... 691
2146 GAACCTATATCTCTCTTCCAAACTACGCGCTTGTGCTGTCTTA 2195
692SerAlaGlnAlaCysTyrAsnG 699
2196 TGTTCCTACAGAAATCTCTGTTCTTTTCAGAAACCTTGCTACACCC 2245
699 luserIleHisGluLeuLysThrLysTyrArgSerPheSerLysGluGly 715
2246 ATACGATTAACGATCTGAAACCAAGTATACACATATCTACACGTAA 2295
716 PheGlySerTrpHisSerValAlaValSerGlyGluValCysAlaSerII 732
2296 ...GGAACCTGGGGAATGATAGTTTCGCTTGAATTCGGTGAAGAGC 2342
732 eProIleValSerAsnGlySerGlyLeuPheSerSerPheSerIlePheS 749
2343 TCCGATTGCTTAGATGAAGTGCCTATTGGACAGTACATGCCCTTCA 2392
749 erLysLeuGlnGlyPheSerGlyThrGlnAspGlyPheGluGluSerSer 765
2393 TGAATTCGACTTGTCTATCCACATCAGCAAGCTTTTAAAGAACAGCGA 2442
766 GlyIleIleArgSerPheSerAlaSerSerPheArgAsnIleSerLeuPr 782
2443 ACAGAACTCTGTGAATTTGGAAGTAGCCGCTGTGTGAATCTTGCCCTTAC 2492
782 oIleGlyIleThrPheGluLysLysSerGln...LysThrArgThrTyrT 798
2493 TATCGGATCCGATTGTGATTAAGATCAGACTGCCAAGATCCAACTGCA 2542
798 yTyrPheLeuGlyAlaTyrIleGlnAspLeuLysArgAspValGluSer 814
2543 ATCTAACCTCTGGT...TATCTGTGATCTCTTTCGTAAGTAACCCGAC 2589
815 GlyProValIleLeuLeuLysAsnAlaValSerTrpAspAlaProMetAl 831
2590 TGTACGACAAACCTGCGAATTAAGCGGTATCTTGGAAAACCTTCGCGT 2639
831 aAsnLeuAspSerArgAlaTyrMetPheArgLeuThrAsnGlnArgAla 848
2640 GAATTTGGCAGACAGAGCTTATAGTCTCTCGTCGAGAGACATTTTGTCT 2669
848 euHis...ArgLeuGlnThrLeuLeuAsnValSerCysValLeuArgGly 863
2690 TTAACCTCAAAATTTGAACCTTTAGCCCAATTTCTTTTGAATTCGCGTCC 2739
864 GluSerHisSerTyrSerLeuAspLeuGlyThrThrTyrArgPhe 878
2740 TCACTCTGCAATTAACAATGTAGACTTAGAGCAAAATACCAATTC 2784

1018 GTTATTTTCTAACAAACCTCGCAATGCTACTCTTAAGAGCGGCA 1067
1068 TTTTATTTTCTAACAAACCTCGCAATGCTACTCTTAAGAGCGGCA 1067
295 LeuAlaIleuSerGlySerLeuSerIleuValIleuGlySer 311
1068 TTTTATTTTCTAACAAACCTCGCAATGCTACTCTTAAGAGCGGCA 1067
1068 TTTTATTTTCTAACAAACCTCGCAATGCTACTCTTAAGAGCGGCA 1067
312 ValLeuPheGlnAsnSerGlnArgPhe.....SerAspGlnGlyLe 326
1118 ATCATTTTCAGAGGCATACTACAGAGCATACAGAGAGTCTCGAGAGT 1167
326 uValArgAsnAlaIleuLeuGlnLysAspAlaIleuSerLeu 343
1168 GACCAAGAAAGCTATAGATCTTCATCGAAATGCTTAATCTCC 1217
343 LuAlaArgAsnGlyAsp...IleLeuPhePheAspProIleValGlnGlu 358
1218 GAGCGACTCGGGAAATAAAGTATTTTCATACATCTTCATC..... 1258
359 SerSerSerLysGlnSerProLeuProSerSerLeuGlnAlaSerValTh 375
1258 1258
375 rSerProThrProAlaThrAlaSerProLeuValIleGlnThrSerAla 392
1259ACGAGCTCAGAGAGCTACATACATACGCTCTCTTGA 1292
392 snArgSer.....ValIlePhe 397
1293 ATTAAGCTGACGAGATCTCGAAATACCTATGAGCGCTACATCGTTTC 1342
398 SerSerGlnArgLeuSerGlnGluLysThrProAspAsnIleuTh 413
1343 TCTGAGAGAACTCTCAGAGAGAGACTTAAGAACTCTGACATCTGAA 1392
413 rSerGlnLeuGlnProIleGlnLeuLysSerGlyArgLeuValLeu 430
1393 GTCATCATTTACAGAGCTCTAGAGCTCTGAGCTCTTACTATTGA 1442
430 yAspArgAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAla 446
1443 AAGATGAGAGTACTGTAGTCCAAATCTATACAGCGAGCTCAGAGATCG 1492
447 LeuLeuIleMetGlnAlaGlyThrSerLeuLysThrSerSerAsp...Le 462
1493 AAGTCTGATGATGAGAGAGAGACTTTTGAAGAGAGCGCTGAGGGGT 1542
462 uLysLeuAlaThrLeuSerIleProLeuHisSerLeuAsp...ThrGlu 478
1543 CACTCTCAATGCGCTACCATTAATAGATCTTACATGAGCAATA 1592
478 ySerVal.....ThrIleHisAlaProAsnLeuSerIleGlnLys 491
1593 AAGCTATCATTAAGCGAGCGAGAGATAGATGTTGCTTATACAGCG 1642
492 IlePheLeuSerAsnSerGlyAspGlnAsnPheGlyGlnAsnValGluLe 508
1643 CCTATCATGCTTGTAGATGCTCAGAGAGACTATATGATCATCAATCT 1692
508 uLeuSerLysGlnGlnAsnAsnIleProLeuLeuThrLeuSerLysGlu 524
1693 C.....AGTAAACAGAGAGCTTCTTCAATAGAGCTTCTCCACACA 1736
525GlnSerHisLeuHisLeuProAspGly.....AsnLeuSer 536
1737 GAAGCATGACTACTACATATCCCGCATACCCCAATTTCTAATATCTACG 1786
537 SerHisPheGlyTyGlnGlyAspTrpThrPheSerTrpLysAspSerAs 553
1787 AATCATCTATGGTATCAAGAGAGAGATATTTGTTGGTGCAGAGATGC 1836
553 pGluGlyHisSerLeuIleAla.....AsnTrpThrProLysAsnTyv 568

1837 AACTGCAAAAAAANAAGTCTACCTTAACCTTGACATTAAGAGATACA 1886
568 aProHisProGlnArgIleuSerThrLeuValAlaAsnThrLeuTrpAsn 584
1887 AGCGGAATCCAGAAAGCTCAGAGAGCTTTGGTCTCTAATAGCTGTGGGT 1936
585 ThrTrpSerAspMetGlnAlaValGlnSerMetIleAsnThrIleAla 601
1937 TCTTTTCATGATGCTCGGCTCATTCAGAGGCTCATGAGCGGAGACACA 1986
601 sGlyGlnAlaValLeuPheGlyThrTrpGlySerAlaValSerAsnLeu 618
1987 TTTTATTTTCTAACAAACCTCGCAATGCTACTCTTAAGAGCGGCA 2033
2034 TTTTGCATGAAAGATCAGAAAGAAACCA...CGTACTTATCGTCAATCT 2080
618 hetyrAlaHisAspSerSerGlyLysProIleAspAsnTrpHisArg 634
2034 TTTTGCATGAAAGATCAGAAAGAAACCA...CGTACTTATCGTCAATCT 2080
635 SerLeuGlyTyLeuPheGlyIleSerThrHisSerLeuAspHisSe 651
2081 AGCGGAGGTTATGCTATGAGAGAGAGATTTCTACGCGCTTCTGAATTT 2130
651 rPheCysLeuAlaIleGlnLeuLeuGlyLysSerSerAspPheI 668
2131 CTATTAATTTTCTAACAAACCTCGCAATGCTACTCTTAAGAGCGGCA 2180
668 leThrSerThrGluThrThrSerTrpIleAlaThrValGlnAlaGln... 683
2181 TGCTTAAGAAACATACCATGATATATGAGAGGAGCAATGATACCGAC 2230
684LeuAlaThrSerLeuMetLysIle..... 691
2231 CTGAGAGAGTATAGAGCGCTGATAGATTTTTCAGGAATTTGATCTC 2280
692SerAlaGlnAlaCysTyArgLysGlnIleHisG 703
2281 CCTACCTTTTCTAACAAACCTCGCAATGCTACTCTTAAGAGCGGCA 2330
703 LuLeuLysThrLysTyArgSerPheSerLysGlnGlyPheGlySerTrp 719
2331 ACATGACCACAAATACACTGCTATTCTCTGTTAAG...GGAAGCTGG 2377
720 HisSerValAlaValSerGlyValCysAlaSerIleProIleValSe 736
2378 GGAATGATGCTGCTGATAGATGATGAGAGACTATCCGGATGTTGC 2427
736 rAsnGly...SerGlyLeuPheSerSerPheSerIlePheSerLysLeu 752
2428 TTCAGAGAGTGGTCTGAGGATACCCACAGCCATTTCTAAACCTAG 2477
752 LngLysSerGlyThrGlnAspGlyPheGlnGlnSerSerGlyGluLe 768
2478 AGATGATCTATGACATCAGATACCTTTAAGAAACGAGCAGAAAGCC 2527
769 ArgSerPheSerAlaSerSerPheArgAsnIleSerLeuProIleGly 785
2528 CGTCTTCTTCCAAAGTGAAGACCTTCAATCTACAGGTTCTCTAGAGGAT 2577
785 eThrPheGlnLysLysSerGlnLysThrArgTrpTyTrpPheLeu 802
2578 AAAATTTGAGAAATTTCTCGAATAGTCT...ACGATATCTCTCATTA 2633
802 LuAlaTyrIleGlnAspLeuLysArgAspValGlnSerGlyProValAl 818
2634 ..GCTTACGTTCCGAGATGATTTGCTAATGATCCAGCGTGCAGACAACT 2671
819 LeuLeuLysAsnAlaValSerTrpAspAlaProMetAlaAsnLeuAspSe 835
2672 CTATGATGTTCTGAGGATTTCTTGTGCGACATGATGATCAAGCTGTCTAG 2721
835 rArgAlaTyMetPheArgLeuThrAsnGlnArgAlaLeu...HisArgL 851
2722 ACAAGCTCTTCTTACGTCCTGGAATCATCATGCTTCTTCTTCAACT 2771

203 yclValaIle...CysCysSerAsnLeuIleCysSerGluAsnValAsp 219
 98880 TCGAGTAAATTAATACACAGAACTTCAATTAACAGGAATACCTCT 98928
 219 rleuPhePheThrGluAsnSerAlaThr...AsnGlyGly 231
 98929 ..CTTGATTTTCTGAATAATAGTGTGACAGCGACCGACGAATAGAGGA 98976
 232 AlaIleCysCysIleSerAspLeuAsnThrSerGluGlySerLeu 248
 98977 GCTCTT.....TCTGGAGATCCGAGTCTTAC 99002
 248 rleuAlaCysAsnGluThrLeuPheAlaSerAsnSerAlaGlyGlu 265
 99003 CATATCTGGAGTACAGTGTACTTTCACAGAAACCACTGATGTA 99052
 265 ysglyGlyAlaIleThrAlaLysHisMeValLeuArgTyr....Asn 279
 99053 ATGGCGGAGCCATTATGCTAAGAGCTTACACTGGC..TTCGGGGGGG 99101
 280 GlyProValSerPheIleAsnAsn.....SerAlaIle 291
 99102 GGGGATATCTCTTTCTTACAAATATAGTCCAAAGTACCACCTGCA 99151
 291 eglGlyAlaIleAlaIleGlnSerGlyGlySerLeuSerIleLeuAlaG 308
 99152 TCGTCGACCACTTTCTATACCTGCGACGACGACGACGACGACG 99201
 308 yglGlyGlySerValLeuPheGlnAsnAsnSer...GlnArgThrSer 323
 99202 AAGCAGGGGACATTACTTCAATGGGAATGCCATTGTGCACTACAC 99251
 324 GlnGlyLeuValArgAsnAlaIleTyrLeuGluLysAspAlaIle 340
 99252 CAACGTACAAAGAAATTTCTTACATAGATCTACGCAAAATCAG 99301
 340 rSerLeuGluAlaArgAsnGly...AspIleLeuPheAspProIle 356
 99302 GAATTTACGTGCAATATCTCCGCAATACATCTTTTCTACATCC 99351
 356 aGlnGluSerSerSerGlyGluSerProLeuProSerSerLeuGln 372
 99352 CTGCTAATACGGCTCGGATTTCTACAGATCTTAAATCTCAATAG 99401
 373 SerValThrSerProThrProAlaThrAlaSerProLeuValIle 389
 99402 GATCGAGGTAAT.....AGTAC 99418
 389 rSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerGlu 406
 99419 AGATTATAGTGGGTGATTTCTTTCTGCGCAAAAGCTCTGCAAGAT 99468
 406 Lu...LysThrProAspAsnLeuThrSerGlnLeuGlnGlnProIle 421
 99469 AAGCAAAAGTTCAGACAACTCTTACCTGACGTAAGAGCTGTACT 99518
 422 LeuLysSerGlyArgLeuValLeuLysAspArgAlaValIleSer 438
 99519 CTAACTGCGAGAAATTTAGTACTTAAAGTGTGTCTGACTGTGAT 99568
 438 oSerLeuSerGlnAspProGlnAlaLeuLeuIleMetGluAlaGly 455
 99569 ACGCTTTACTGACCGCGGTTCTCTGTTATTTATGATCCGCGACA 99618
 455 eLeuLeuThrSerSer...AspLeuLysLeuAlaThrLeuSerIle 470
 99619 CGTTAAAGCAAGTACAGAGGAGTCACTTAAACAGGCTTTCCATCT 99668
 471 LeuHisSerLeuAspThrGluLysSerValThrIleHisAlaPro 485
 99669 GTAGACCTCTTTAGCGGAGGTAAGAAATGTGAATTCGCTGCTGCA 99718

486AsnLeuSerIleGlnLysIlePheLeuSerAsnSer 498
 99719 AAGTAAATATGACCCCTTACAGTGTCCGATCTCTCTTTTGCATACCA 99768
 498 yAspGluAsnPheThrGluAsnValGluLeuLeuSerLysGluAsn 514
 99769 GG.....AATGCTTATGAAATACAGAC...TTAGGAAAACTCAAG 99809
 515AsnIleProLeuLeuThrLeuSerLysGluGlnSerIle 527
 99810 TTTTCATTTTGTCACCTCTCTCTCTGCTTACCAACACTACAGATGT 99859
 527 sLeuHisLeuProAspGluAsnLeuSerSerHisPheGlyTyrGln 544
 99860 TCCACCGCTTCTTACAGTACAGCAATCTTACGCACTATGGGTACA 99909
 544 sPThrThrPheSerThrPylAsp.....SerAspGluGlnHis 556
 99910 CTGGGGAATGCTTGGGTGTGATACCGCAAGCACTCAAAAGCTAAG 99959
 557 SerLeuIleAlaAsnThrProLysAsnTyrValProHisProGlu 573
 99960 ACAGGACATTAAGCTTGGACCAATACAGCTTCTCCGAATCCCTGAC 100009
 573 glnSerThrLeuValAlaAsnThrLeuThrAsnThrTyrSerAsp 590
 100010 TCAACGACCTTTAGTCCCAATATACCTTTGGGATCTTTTCACGACAT 100059
 590 InAlaValGlnSerMetIleAsnThrIleAlaHisGlyAlaTyrLeu 606
 100060 AAGCATTCAAAGGTGCTATAGAGAGAGTGTCTTACCTTTGTCAGAT 100109
 607 PheGlyThrThrPheGlySerAlaValSerAsnLeuPheTyrAlaHis 623
 100110 CAGGCTTCTGGCTGGCGGAGTCCGCAAT...TTCCTAAGATAAGAT 100156
 623 rSerGlyLysProIleAspAsnThrHisIleArgSerLeuGlyTyr 640
 100157 GAAAGCGGAAAA...CCCAATACCGTCATAAATCGTGATGTGCTA 100203
 640 heGlyIleSerThrHisSerLeuAspAspHisSerPheCysLeuAla 656
 100204 TCGGAGGTGACGCGCAACTGTCTGAAAACTTATTAAGTCTTCCCT 100253
 657 GlyGlnLeuGlyLysSerSerAspSerPheIleThrSerThrGlu 673
 100254 TCCCAACTCTTTCGTAAGCATTAACATTTCTTACGCGCTAAATCAT 100303
 673 rThrSerThrIleAlaThrValGlnAlaGlnLeuAlaThrSerLeu 688
 100304 TCATACATATGAGAGAGCTTATATCCAAACACTTACAGATGTAGT 100353
 688 688
 100354 GGTTCATAGGTGTCTCTTACATAAATCTCGGCTTGGAGTCATAAA 100403
 689 ...MetLysIleSerAlaGlnAlaCysTyrAsnGluSerIleHisGlu 704
 100404 CCCCTGTTTAAAGGCGACGCTCTTATACCCAGCTAGTAATAGATCT 100453
 704 uLysThrLysTyrArgSerPheSerLysGluGlyPheGlySerThr 721
 100454 GAAGCAAAAGTATACGCTAT...CTGAGGTGAAGAGTCTTGAGGGA 100500
 721 eValAlaValSerGlyValAlaCysAlaSerIleProIleValSer 737
 100501 ATATGCTTTTAACTATGATGTGGAGCTTCTTCATTTATCTTATC 100550
 738 GlySerGlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGly 754
 100551 TACCGCATTTGTTGATACCTATGCTCATACATCAAACTGAACTGAC 100600
 754 eSerGlyThrGlnAspGlyPheGluGluSerSerGlyGluIleArg 771

223 ThrGlyAsnSerAlaThr.....As 229
|||.....
833 ATTGACATTAACCTCAGACGAGCAGCTCCTCACACACCGGATATGTC 882
229 nGlyAlaIleCysCysIleSerAspLeuAnThrSerGlyGly 246
|||.....
883 AGGAGTCTATCTGCT.....TATAACTAGTACGATCTA 923
246 erLeuSerLeuAlaCysAsnGlnIleThrLeuPheAlaSerAsnSerAla 262
|||.....
924 AGGTACCTCCTCAGTCAATCAGATGTTACTCTTACGACACATATCATCG 973
263 LysGlyLysGlyAlaIleTyrAlaLysHisMetValLeuArgTyrAs 279
|||.....
974 ACAACAGCGGAGGAGCATCTATCTGAAAAAGCTCGAAGCTGGCTCCGG 1023
279 nGlyProValSerPheIleAsnAsnSer.....AlaLysI 291
|||.....
1024 AGGACTTACCTATTCAGTAAATAGTGTCAATGAGGATACAGCTCTCA 1073
291 IeGlyAlaIleAlaIleGlnSerGlySerLeuSerIleLeuAla 307
|||.....
1074 AAGGTGGAGCCTAGCTATCGAAGATAGTGGGAATGAGTTATCCGCC 1123
308 GlyGlyLysSerValLeuPheGlnAsnAsnSerGlnArgThrSerAspI 324
|||.....
1124 CATACTGCTGACATTCCTCTTTTACGCAATACAGTCACCTTACTACTCC 1173
324 nGlyLeuValArgAsnAlaIleTyrLeuGlyLysAspAlaIleLeuSer 341
|||.....
1174 TGGGACGATAGAGTATGATGATGAGAACGAGTCAAGATGACAG 1223
341 erLeuGlnAlaArgAsnGly...AspIleLeuPhePheAspProIleVal 356
|||.....
1224 CTTCGCTTCTGCTGCTGTAGAGCCTACTCTTATGATCCCTACT 1273
357 GlnGlySerSerSer.....LysGlySerPr 365
|||.....
1274 ACAGATCATCTCACACACAGTACAGATGCTTTAAAGTTAAACAGACCTC 1323
365 OLeuProSerSerLeuGlnAlaSerValThrSerProThrProAlaThr 382
|||.....
1324 GGCAGATTCCTGACACTACAAATATACA..... 1348
382 IAserProLeuValIleGlnIleThrSerAlaAsnArgSerValIlePheSer 398
|||.....
1349GGGACATCATCTTCCACA 1366
399 SerGluArgLeuSerGlyGlyLysThrProAsp...AsnLeuThrSe 414
|||.....
1367 CGAGAAAGTTATCGACACAGACAGCCGCCACATTCCTAAATCTTACTTC 1416
414 rGlnLeuGlnIleProIleGlyLeuLysSerGlyArgLeuValLeuLys 431
|||.....
1417 GAAGCTACTACAGCCTTACTCTTTCAGAGGACTCTATCTTAAAC 1466
431 spArgAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeu 447
|||.....
1467 ATGGAGTACTCTGAGACTCAGGACTTACCTCAACAGCAGACTCTGCT 1516
448 LeuIleMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLe 464
|||.....
1517 CTCGAAATGAGACGTAGCAACTACTCTACACCTGCTATACAGACCAT 1566
464 uAlaThrLeuSerIleProLeuHisSerLeuAspThrGlyLysSerValT 481
|||.....
1567 AAACAATTGTGTCATTACATCAGTCTATAGACGCTGCAAAAGAGCGCA 1616
481 hrIleHis.....AlaProAsnLeuSerIleGlnLysIlePhe 493
|||.....
1617 AAATGAAACCAAGCTACGTCAAAAAATCTGACTTTATCTGGAACATC 1666

494 LeuSerAsnSerGlyAspGluAsnPheTyrGluAsnValGluLeuLeu 510
|||.....
1667 ACTTATTCGACCCGACCGGACGCTTTTATGAAATTCATAGTTTAAAGAA 1716
510 rLysGlnGlnAsnAsnIleProLeuLeuThrLeuSer...LysGlnGln 526
|||.....
1717 TCCTCAGTCTCAGACATCTTAGAGCTCAAAAGCTCTGGAAGTGAACA 1766
526 erHisLeuHisLeuProAspGlyAsnLeuSer.....HisPheGly 540
|||.....
1767 GCACCCGAGTCACTCCAGATCTATATGCGTGAAGAAATTCATACCGC 1816
541 TyrGlnGlyAspTrp...ThrPheSerTrpLysAspSerAspGlyHis 556
|||.....
1817 TATCAGGAGACTTGGGCCCAATCTTTG.....GGACAGGGGC 1857
556 sSerLeuIleAla.....AsnTrpThrProLysAsnTyrValProHisP 571
|||.....
1858 TTCTACGACTGCAACCTCACTCACTGATAAAGCTGATATCTCTATC 1907
571 rGluArgGlnSerThrLeuValAlaAsnThrLeuTrpAsnThrTyrSer 587
|||.....
1908 CCGAGCGTATCGGCTTACTCCCTAATAGCTATGGAATGCAATTTATA 1957
588 AspMetGlnAlaValGlnSerMetIleAsnThrIleAlaHisGly..... 602
|||.....
1958 CATATTACTCTCTCATTTACTTATGAGACTGCAAAAGAGGTTGCA 2007
603 GlyAlaTyrLeuPheGlyThrTrpGlySerAlaValSerAsnLeuPhe 619
|||.....
2008 GGGAGACCGTCTCTT.....TGCTGTGCGGATATCTACTCTCTC 2050
619 YrAlaHisAspSerSerGlyLysProIleAspAsnTrpHisArgSer 635
|||.....
2051CATAGGATAGTACAAAAACACGACCGCGCTTGGCATTGGACT 2095
636 LeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAsnHisSerPh 652
|||.....
2096 GCGGCTTATGTCATAGAGGAAGAACTACATCTTCTAGATTAAGATTCT 2145
652 eCysLeuAlaIleGlyGlnLeuLeuGlyLysSerSerAspSerPheIle 669
|||.....
2146 TAGTCTCATTTGTCTGCTCTTGAAGATGATGAGACTACTTGTAG 2195
669 hrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaGlnLeuAla 685
|||.....
2196 CTAAAGATCAAGGTACAGCTACGAGGAACTCTATTTACACGACAC 2245
686 ThrSerLeuMetLysIle.....SerAlaGlnAlaCysTyrAsn 691
|||.....
2246 CAACCTTATCTCTCTCTCTTCCAACTACGCCCTTGTCTGTCTTA 2295
692SerAlaGlnAlaCysTyrAsn 699
2296 TGTTCCTCAGAGATTCCTGTTCTCTTTAGAAACCTTAGCTACACCC 2345
699 IuSerIleHisGluLeuLysThrLysTyrArgSerPheSerLysGlyGly 715
|||.....
2346 ATACGATTAACGATCTGAAACCAAGATTAACAACATATCACTGTTAAA 2395
716 PheGlySerTrpHisSerValAlaValSerGlyGluValLysAlaSerI 732
|||.....
2396 ...GGAACCTGGGGAATGATAGTTTCCCTTTAGAAATTCGCGTGAAGC 2442
732 eProIleValSerAsnGlySerGlyLeuPheSerSerPheSerIlePhe 749
|||.....
2443 TCCGATTGTGTTAGTAAAGTCTCTATTGTAGACAGTATGCCCTTCA 2492
749 erLysLeuGlnGlyPheSerGlyThrGlnAspGlyPheGlyLysSer 765
|||.....
2493 TGAATTTCAGTGTGTCTATGACATGCAAGGTTTAAAGAACAGGGA 2542
766 GlyLuiLeuArgSerPheSerAlaSerSerPheArgAsnIleSerLeuP 782

189 ValSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnGlyAla 205
189 ValSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnGlyAla 205
643 ACCGGCCAGTAGCTTATGATATATAAGAGAACTGGGGTGGGC 692
205 aIleCysCysSerAsnLeuIleCysSerGlyAsnVal.....Asp 219
693 TCTC.....GGCTTCAAGCCACCTCTCATTAACCAAAATAGCT 733
219 roleuPhePheThrGlyAsnSerAlaThrAsn.....GlyGly 231
734 CCTTTTCTCTCTGGAAACACTGCACACAGATGCTGCAGCAAGGGCGG 783
232 AlAlIleCysCysIleSerAspLeuAsnThrSerGlySerLeuSe 248
784 GCCATTATATGT.....GAAAAACAGAGAGAGACTCTACTCTTAC 824
248 rLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaGly 265
825 TATCTCTGGAAATAAACTCTGACCTCCCGAAGACCTCTCACTAAC 874
265 ySGlyGlyAlaIleTyrAlaLysHisMetValLeuArgTyrAsnGly 281
875 AAGGGGAGCATCTGTGCTCCATGTCTAGATCTTCCGCTGTGGCCT 924
282 ValSerPheIleAsnAsn.....SerAlaLysIleGly 293
925 ACCCATTTTCAATATATAGATGCGGAGACACACTGCAAGCAAGGGCG 974
293 yAlaIleAlaIleGlnSerGlyGlySerLeuSerIleLeuAlaGly 310
975 CCTATTGCAATTCGCCACTCTGCATCTTAACTCTCTCAAAATYCAAG 1024
310 ySerValLeuPheGlnAsnAsnSer...GlnArgThrSerAspGln 325
1025 GAGCATACACTCTCTGGCAACACTTAACCTCAACCTCCGCGCAACA 1074
326 LeuValArgAsnAlaIleTyrLeuGlnLysAspAlaIleLeuSer 342
1075 TCGACACGAGATGCTATCTACTGCGATCGTCACCAAAATACCACTT 1124
342 uGluAlaArgAsnGly...AspIleLeuPhePheAspProIleVal 358
1125 AAGGGCCACCCCAAGCCCAATCTATCTTATGATCCGATTCATCTTA 1174
358 luserSerLysGlySerProLeuProSerSerLeuGlnAlaSerVal 374
1175 ACACCAAGAGCTTCAGACGTTCTG..... 1200
375 ThrSerProThrProAlaThrAlaSerProLeuValIleGlnThr 391
1201 ACCATCAACCAACCGATAGCAACTCGCCTTA.....GATTA 1238
391 AsnArgSerValIlePheSerSerGluArgLeuSer...GluGluGlu 407
1239 TTCAGCAAGCATTTTATTTTGGGAAAGCTCTCTGCAGATYGAAGCA 1288
407 ySThrProAspAsnLeuThrSerGlnLeuGlnProIleGluLeuLys 423
1289 AAGCTGCTGATTACTTACATATATATTAACCAACCATTTGCTTACGC 1338
424 SerGlyArgLeuValLeuLysAspArgAlaValLeuSerAlaProSe 440
1339 TCTGGACCTTACACTCAAGAAATGCGATGATGATGCAATGCGTTT 1388
440 uSerGlnAspProGlnAlaLeuLeuIleMetGluAlaGlyThrSer 457
1389 CACAGAGACTGAAGCTCTACACTCTCTCAACACGAGCAACAAGCTCA 1438
457 ySThrSerSerAsp...LeuLysLeuAlaThrLeuSerIleProLeu 472
1439 AAGCAATATCTAAGCTATCACTTACCAAACTGTCTGATCTTCTTCT 1488
473 SerLeuAspThrCluLysSerValThrIleHisAlaPro..... 485

1489 GCCCTPACAGGAATAAGTGTGTCATTTGAAACACAGCAACCAACA 1538
486AsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAsp 500
1539 AACTATTAACCTTAACCTCTCTCTTCTTTCAGATAGTAGGCG.... 1584
500 LysAsnPheTyrGluAsnValGluLeuLeuSerLysGluGlnAsn 516
1585 ..AATTTTATGAAGC...CATACGATAAACCAAGCTTCAGCAGACCT 1629
516 e.ProLeuLeuThrLeuSerLysGluGlnSerHisLeuHisLeuPro 532
1630 TTGCTGTATATAGCTGCTACTGCTCTAGCCGATATATATATC... 1676
533 GlyAsnLeuSer.....HisPheGlyTyrGly 542
1677 GCGCTTCTACCTCTCCAGTACAACTCCAGACCTCATTTACGGCTATCA 1726
542 nGlyAspThrPheSerThrLysAspSerAspGluLysIleSerLeu 559
1727 GGCATTTGGAAAGCACTTGGCCAGACATCACTCAAAATCAGGAA 1776
559 lAlaAsnThrPheProLysAsnTyrAlaProHisProGluArgGln 575
1777 CTATGCTTGGCTAAGTACGAGCTACCACTTAATCTGACCGTAGAGCT 1826
576 ThrLeuValAlaAsnThrLeuThrPheAsnThrTyrSerAspMet 592
1827 TCCGTACTTCCCGATTCATTAATGCGCATCTTATACATTCCTCCTCT 1876
592 lGlnSerMetIleAsnThrIleAlaHisGlyAlaTyrLeuPheGly 609
1877 ACACAGATCATGACATCTCAAGCAATAGTATCTATCAGCAAGAGAG 1926
609 hTTPGlySerAlaValSerAsnLeuPheTyrAlaHisAspSerSer 625
1927 TCTGCGCATCAAGAACTCGCAATTTCTTCAT...AAGCAATAATCG 1973
626 LysProIleAspAsnThrPheHisArgSerLeuGlyTyrLeuPhe 642
1974 ACTAAC...CAAGCATTCGCAATATAAAGCTACGCTATATTTGAGG 2020
642 eSerThrHisSerLeuAspAspHisSerPheCysLeuAlaIleGln 659
2021 AAGTCTGATGATTTTCTGAAATATCTTCAGTACCTTTCTGCAGC 2070
659 euleGlyLysSerSerAspSerPheIleThrSerThrGluThrThr 675
2071 TCTTCGCTAAAGATTAAGACCTGTTATATGTAATAACCTTCATAC 2120
676 TyrIleAlaThrValGlnAlaGln..... 683
2121 TATTTACCGCTGCTATACCTGCAACATCGACATTCCTAGAGAGACT 2170
684LeuAlaThrSerLeuMetLys..... 690
2171 CATGCCCTCATTTGGAACTATCACCGACATGCTGAAAGATATTCCT 2220
691 ..IleSerAlaGlnAlaCysTyrAsnGluSerIleHisGluLeuLys 706
2221 TTTTGAAGCCAGCTTAAGCTACACTCACTAAAGAAATGATGATCT 2270
707 LysTyrArgSerPheSerLysGluGlyPheGlySerThrPheSerVal 723
2271 CGCTTACTTCTTAT...CTGAAAGCTTAAGGCTTTTGACCAATATCT 2317
723 aValSerGlyLysValCysAlaSerIleProIle...ValSerAsnGly 739
2318 TGGGCTCTAGAGCTCGAGAGATCTCTGCTATATCTCCCTAAAGAG 2367
739 eGlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGlyPheSer 755

[illegible]

1633	GTGGATTTCACTGTCGTCTACTGCTGCTAGAGATATTATATC..GATCG	1679
533	YASnLeuSerSer.....HisPheGlyTyrGln	543
1680	CGTTTCACACTCTCCGCAATCCAAACCTCCAGAACCTTATTACGGGTATCAGG	1729
543	LYAspTrpThrPheSerTrpLysAspSerAspGluGlyHisSerLeuIle	559
1730	GACATTGGGAAGACCACTTGGCGAGACATCAAGTCAAAATACAGAACT	1779
560	AlAsnTrpThrProLysAsnTyrAlaProHisProGluAlaGlnSerThr	576
1780	ATGACTTTGGGTAACTACGGGGCTTAAACCCCTAATCTCGAGCGTAGCGCTC	1829
576	rLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMetGlnAlaValG	593
1830	CGTACTTCCCGATCTCTTATGGCGATCCCTTATCTACACTCCGACTATAC	1879
593	InsMetIleAsnThrIleAlaHisGlyAlaTyrLeuPheGlyThr	609
1880	AGCAGATCTGACATCTCAACCGAAATAGATCTACAGCAACGAGACTC	1929
610	TrpGlySerAlaValSerAsnLeuPheTyrAlaHisAspSerSerGlyI	626
1930	TGGGCACTACGGAACTCCGAATTTCTTCCAT...AAGGATTAATCGAGAAC	1976
626	sProlLeaAspAsnTrpHisIAspSerLeuGlyTyrLeuPheGlyIle	643
1977	TAAAC...CAAGCATTCGCCGATTAATAACGTACGGCTATATTGTTGGAGGA	2023
643	eThrHisSerLeuAspAsnHisSerPheCysLeuAlaAlaGlyIleu	659
2024	GTCCGACAGATTTTCTGAAATATCTTCACGTGATCGCTTCGCCAGCTC	2073
660	LeuGlyLysSerSerAspSerPheIleThrSerThrGluThrThrSerTy	676
2074	TTGCGTAAAGATTAATAACCTGTTTATGTTGAAATACCTCTCATATAC	2123
676	rIleAlaThrValGlnAlaGln.....	683
2124	TTTACCGTGGCTATACCTGCAACATCGACATTCCTAGAGAGACTTCCCA	2173
684LeuAlaThrSerLeuMetLys.....	690
2174	TGCCCTCATTTGGAGAGTATACCCGACATCGTGAAGATATTCCTCATTT	2223
691	HisSerAlaGlnAlaCysTyrTrasnGlnSerIleHisGluLeuLysThrLy	707
2224	TTGAAATGCCCATCTAAGCTACGACCTACATAAAATGATATGATGATCTCG	2273
707	sTyrTrpSerPheSerLysGluGlyPheGlySerThrPheHisSerValAla	724
2274	CTATACCTCTATA...CCGTAAGCTCAAGGTCTTGGACCAATATTTGTG	2320
724	AlSerGlyGluValCysAlaSerIleProIle...ValSerAsnGlySer	739
2321	GGCGCTACAGCCGAGAGATCTCGGCTATATCTCCCTAAAGAACCA	2370
740	GlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGlyPheSerG	756
2371	CGCTTTCTTCAGAGGATATTTCCTCTTAAAGTCCAGACAGTCTACAG	2420
756	YnThrGlnAspGlyPheGluGlnSerSerGlyuIleLeuArgSerPheSerA	773
2421	CCGCCAACAAACTTTAAAGAGAGAGTGGCGTGAACCCGCTTTTGATG	2470
773	lAspSerPheArgAsnIleSerLeuProIleGlyIleThrPheGluLys	789
2471	ATGAGAACCTAGTGAACCTGCTTATCCCTGCTGGCATTCGGTACAAAAA	2520
790	LysSerGlnLysThrArgThrTyrTyrTrpPheLeuGlyAlaTyrIleG	806
5521	ATCTCGAAGATGAAAAATATATTTCGAGATTCTCTGACCACTATGG	2570

189 ValSerProIleSerPheAlaIArgAsnArgAlaAspLeuAsnGlyVal 205
 743 ACCGGCCAGTTAGTCTTATGATATAAAGGAAAGTGGGGGTGGGC 792
 205 aIlecySerAsnLeuIlecySerGlyAsnVal.....Asn 219
 793 YCTC.....GGCTTGAAGCCACCTCCCTCATTTACTCAAAATACCT 833
 219 rOleuPhePheThrGlyAsnSerAlaThrAsn.....GlyGly 231
 834 CCTTTCTCTCTGTGAACACTGCACAGATGTGCAGGCAAGGCGGG 883
 232 AlaIleCyScyIleSerAspLeuAsnThrSerGlyGlySerLeuSe 248
 884 GCCATTATTGT.....GAAAAACAGAGAGACTCTACTACTTAC 924
 248 rLeuAlaCyAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysGlu 265
 925 TATCTCTGGAATAAAGCTTCACCTTCGCCAGACACTTCACGTAATC 974
 265 ysglyGlyAlaIleTyrAlaLysHisMetValLeuArgTyrAsnGlyPro 281
 975 AAGCGGAGCATCTGTGCCATGTGTATGATCTTCCGCTCGCCCT 1024
 282 ValSerPheIleAsnAsn.....SerAlaLysIleGly 293
 1025 ACCCTATTTTCAATAATAGATGCGGGAACACAGCTGCAGCGAGGCGG 1074
 293 yAlaIleAlaIleGlnSerGlyGlySerLeuSerIleLeuAlaGlyGln 310
 1075 CCGATTTCATTTGCCGACCTCGACCTGATTTAAGTCTCTGCAAAATCAG 1124
 310 LysSerValLeuPheGlnAsnAsnSer...GlnArgThrSerAspGlnGly 325
 1125 GAGCATCATCTCTCTGCAACACTCAACCTCAACCTCCGCGCCACA 1174
 326 LeuValArgAsnAlaIleTyrLeuGluLysAspAlaIleLeuSerSer 342
 1175 TCGACGCGGAATGCTATCTACTGCGATCGTACGAAATAATTACGACTT 1224
 342 uGluAlaIArgAsnGly...AspIleLeuPhePheAspProIleValGln 358
 1225 AAGCGCACCACCAAGGCCAATCATCTATTCTCATATCCGATTCGATCTA 1274
 358 LysSerSerLysGluSerProLeuProSerSerLeuGlnAlaSerVal 374
 1275 ACACCCAGGAGCTTCAGACGCTCTG..... 1300
 375 ThrSerProThrProAlaThrAlaSerProLeuValIleGlnThrSerAl 391
 1301 ACCATCAACCAACGATAGCAACTCGCCTTA.....GATTA 1338
 391 AsnArgSerValIlePheSerSerGluArgLeuSer...GluGluGlu 407
 1339 TTCACGACGACGATTTCTCTGGGAAAGCTCTCTGCACATGCAAGCA 1388
 407 ySThrProAspAsnLeuThrSerGlnLeuGlnProIleGluLeuLys 423
 1389 AAGTGTGATTAACCTCATCTATATTAAAGCAACCATTTGGCTTACCC 1438
 424 SerGlyArgLeuValLeuLysAspArgAlaValLeuSerAlaProSerLe 440
 1439 TCTGGACCTTAGCACTCAAGGAATGTGAGTTAGTCAATGATGTTT 1488
 440 uSerGlnAspProGlnAlaLeuLeuIleMetGluAlaGlyThrSerLeu 457
 1489 CACACAGACGCTAAGGCTCTACACCTCATGCAACCCAGCAACCAAGCTCA 1538
 457 ySThrSerSerAsp...LeuLysLeuAlaThrLeuSerIleProLeuHis 472
 1539 AAGCAGATACGGAAGCTATCAGCTTACCAAACTGTGTGATCTTCT 1588
 473 SerLeuAspThrGluLysSerValThrIleHisAlaPro..... 485

1589 GCCTTAGAGGAAATAGAGTGTGTCCATTTGAACAGCAGGACCAACAA 1638
 486AsnLeuSerIleGlnLysIlePheLeuSerIleSerGlyAsp 500
 1639 AACTATTAACCTCAACCTCTCTTGTCTTTCAAAGATAGTACGCGC... 1684
 500 LuAsnPheTyrGluAsnValGluLeuLeuSerLysGluGlnAsnIle 516
 1685 ..AATTTTATGAAGAACCATAGATAAACCAAGCTTACCCAGCTTTG 1732
 517 ProLeuLeuThrLeuSerLysGluGlnSerHisLeuHisLeuProAspG 533
 1733 GTCGATTTCACCTGCTGCTACGCTGCTAGCCATATTATATC...GATCC 1779
 533 yAsnLeuSerSer.....HisPheGlyTyrGln 543
 1780 CTTCTCCTCTCTCCAGTACAACTCCAGAACCTCATATTACGGTATCCG 1829
 543 LysAspThrPheSerThrLysAspSerAspGluGlnHisSerLeuIle 559
 1830 GACATTGGGAAGCCACTGGGCGAGACATCACTGCAAAATCAGAACT 1879
 560 AlaAsnThrPthrProLysAsnTyrValProHisProGluArgGlnSer 576
 1880 ATGACTTCGCTACCTACGCGCTACACCCCTAATCTGAGCTAGAGCTTC 1929
 576 rLeuValAlaAsnThrLeuThrPasnThrTyrSerAspMetGlnAlaVal 593
 1930 CGTACTTCCGATTCATATTATGCGATCTTATGACATTCGACACTAC 1979
 593 InSerMetIleAsnThrIleAlaHisGlyAlaTyrLeuPheGlyThr 609
 1980 AGCAGATCATGACATCTCAGCAAGATAGTATCTACGACAGAGAGCTC 2029
 610 TrpGlySerAlaValSerAsnLeuPheTyrAlaHisAspSerSerGly 626
 2030 TGGCATCAGAGAACTGCGAATTTCTTCAT...AAGCTAATATCAGGAC 2076
 626 sProIleAspAsnThrHisArgSerLeuGlyTyrLeuPheGlyIle 643
 2077 TAAC...CAACATTCACACATTAAGCTACGCTATATTCTTGGAGCA 2123
 643 erThrHisSerLeuAspAspHisSerPheCysLeuAlaIleGlnLeu 659
 2124 GTGCTGAAGATTTTCTGAATAATCTCAGTATGAGCTTCTGCGAGCTC 2173
 660 LeuGlyLysSerSerAspSerPheIleThrSerThrGluThrThrSer 676
 2174 TTCGCTAAAGATAAAGACCTGTTATAGTTGAAATACCTCTCATACTA 2223
 676 rIleAlaThrValGlnAlaGln..... 683
 2224 TTTAGCTTCCTATACCTGCAACATCGAGCATTCCTTAGAGAGACTTCCA 2273
 684LeuAlaThrSerLeuMetLys..... 690
 2274 TGCCCTCATTTGGAATATCACCGCATGCTGAAAGATATTCCTCAT 2323
 691 IleSerAlaGlnAlaCysTyrAsnGluSerIleHisGluLeuLysThr 707
 2324 TTTGATGCCAGCTAAGCTACAGCTACACTAAATATATAGTATGACTCG 2373
 707 sTyrArgSerPheSerLysGluGlyPheGlySerThrHisSerValAla 724
 2374 CTATACCTTCCAT...CTGAACTCAAGCTCTTGGACCATATACCTGC 2420
 724 aIleGlyGluValCysAlaSerIleProIle...ValSerAsnGlySer 739
 2421 GGGCTTAGAGCTCGAGAGATCTGCTGCTATATCTCCCAAGAGCA 2470
 740 GlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGlyPheSer 756

```
192 leSerPheAlaIrgAsnArlaAspleuAnGlyAlaIleCysCys 208
      ::::::::::::::::::::
511207 TAAGCTTGGCAAAACAAACGCAACGCAAAAGCGCGTCCCTCATTTCC 511256
      ::::::::::::::::::::
209 SerAsnLeuIleCysSerGlyAsn...ValAsnProLeuPhePheThrG1 224
      ::::::::::::::::::::
511257 ACGGGAGGATTACATTTAACATATACGTTTAACTGACATCATTTTTCGA 511306
      ::::::::::::::::::::
224 YAsnSerAlaThrAsn..... 229
      ::::::::::::::::::::
511307 AAATACCGCGGCGAAACAATGGCGAGCCATTTCACAGCAAGCTAGCACTT 511356
      ::::::::::::::::::::
229 ..... 229
      ::::::::::::::::::::
511357 TTATTACGACCAACAACCAATTAACCTTTATTAACAATATGTTGACCCCA 511406
      ::::::::::::::::::::
230 .....GlyGlyAlaIleCysCysIleSerAspleuAsnThrSe 242
      ::::::::::::::::::::
511407 ACCTGACGTACAGGGGAGGAGCCATTACTGTAGTAGT.....ACATGAGC 511450
      ::::::::::::::::::::
242 rGluIleGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAlas 259
      ::::::::::::::::::::
511451 CCCCACACCACTTAACTACTATTCAGACACCGGCACTCACTTATATAG 511500
      ::::::::::::::::::::
259 eAsnSerAlaIleGlySerGlyAlaIleTyAlaIleHisMetVal 275
      ::::::::::::::::::::
511501 CAAATACACCAATTAATCTGCTGGCGCATTTATATCACAATCTACATG 511550
      ::::::::::::::::::::
276 LeuArgTyTrpAsnGlyProValSerPheIleAsnAsnSer..... 288
      ::::::::::::::::::::
511551 CTTCCTCTGAGACCTGACCTGTTTAAACAACCTGCTGATATGATAC 511600
      ::::::::::::::::::::
289 .....AlaIleGlyAlaIleAlaIleCysSerGlySerLeuS 304
      ::::::::::::::::::::
511601 TGCAGCTCCCTTAGACAGACCAATTCGATTCGACATTCGATCTTTTCA 511650
      ::::::::::::::::::::
304 eTlleLeuAlaGlyGluGlySerValLeuPheGlnAsnAsnSer..... 318
      ::::::::::::::::::::
511651 GTCCTTCGCGCTTCGTCGACACATCACTTTCGACAGCAACACATACATC 511700
      ::::::::::::::::::::
319 ...GlnArgThrSerAspGlnGlyLeuValArlgAsnAlaIleTyLeu.. 333
      ::::::::::::::::::::
511701 AAGAGAGCTTCTGAGTCAAGACCACTACCAAGAAATTTCTATTACATCGG 511750
      ::::::::::::::::::::
334 GlnIleAspAlaIleLeuSerSerLeuGluAlaArlgAsnGlyAsp... 349
      ::::::::::::::::::::
511751 AAACACCAATCTAAGATGTACAGCTCCGACGCTTCACACCATATCTA 511800
      ::::::::::::::::::::
349 leLeuPheAspProIleValGlnGluSerSerLeysGluSerPro 365
      ::::::::::::::::::::
511801 TCTACTTCTATGATCCATTAACA..... 511823
      ::::::::::::::::::::
366 LeuProSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAl 382
      ::::::::::::::::::::
511824 .....ACTGACATCTGACAGCTCTTCAGATGCTTAACCTTAAGG 511867
      ::::::::::::::::::::
382 aSerProLeuValIleGlnThrSerAlaAsnArgSerValIlePheSers 399
      ::::::::::::::::::::
511868 TCCGTGACCTTCGAGGAGATCCCATATCAAGAAACCATTCGATTTTCG 511917
      ::::::::::::::::::::
399 eGlnIleArgLeuSerGluGluGlyThr...ProAspAsnLeuThrSer 414
      ::::::::::::::::::::
511918 GAGAGAACCTCTCGGAGACCAACCTGCAGAAAGCTGATATCTCAAACT 511967
      ::::::::::::::::::::
415 GlnLeuGlnGlnProIleGlnLeuLeuSergIArgLeuValLeuLeuSAs 431
      ::::::::::::::::::::
511968 ACAATTCAGCAACCTTACTCTTCGAGGAGGAGCACTCTCTTTAAATC 512017
      ::::::::::::::::::::
431 pArgAlaValLeuSerAlaProSerLeuSergIAsnProGlnAlaLeuL 448
      ::::::::::::::::::::
512018 AGGAGTCACTCAAGTTCGCAAGTCTTTTCGCAATCTCCGGGCTTACCC 512067
      ::::::::::::::::::::
448 eulleMetGluAlaGlyThrSerLeuSergIThrSerSerAspleuSleu 464
      ::::::::::::::::::::

512068 TCCCTATGATGACGAGGACCATTAGAAAC...GGTGAATGATCATCTA 512114
      ::::::::::::::::::::
465 AlaThrLeuSerIleProLeuHisSerLeuAspThrGluIleSersValTh 481
      ::::::::::::::::::::
512115 TCAATTAATCTTCTCAATGTATGATTCCTTAAAGAGCCAGAAAGATAC 512164
      ::::::::::::::::::::
481 rIleHisAlaProAsnLeuSerIleGlnIlePheLeuSerAsnSer 497
      ::::::::::::::::::::
512165 CCTAAACACCAACACAGCAACT...CAGACAGTCACTTATCTGTGATCC 512211
      ::::::::::::::::::::
498 .....GlyAspGluAsnPheryrGluAsnValGluLeuLeu 509
      ::::::::::::::::::::
512212 TCTCTCTGTAGATCCTTGTGAATGTCTACGAAGATGC..... 512252
      ::::::::::::::::::::
510 SerIleSergluGluAsnAsn.....IleProLeuLeuThrLeuSerly 523
      ::::::::::::::::::::
512253 .....TCTTGAATTAACCTCAAGTCTTTTGTGTCTCACTCTTACTGC 512296
      ::::::::::::::::::::
523 sGluGln...SerHisLeuHisLeuProAspGlyAsnLeuSerSer... 537
      ::::::::::::::::::::
512297 TCACGACCCCGCGAAATATTCACATCAGAC.....TTAGCTCTGATC 512340
      ::::::::::::::::::::
538 .....HisPheGlyTyrglnGlyAspTrpThrPhe 547
      ::::::::::::::::::::
512341 CCCTGAGAAAAATCTATCCATTTGGGATACCAAGGAAATGGGCAATTA 512390
      ::::::::::::::::::::
548 SerTrpIleAspSerAspGluGlyHisSerLeuIleAla.....AsnTr 562
      ::::::::::::::::::::
512391 TCTTGGCAAGGATCTGCGACTAAATCCAAGCAGGACTCTTACTCG 512440
      ::::::::::::::::::::
562 PThrProIleAsnTyValProHisProGluArgIleSerThrLeuValA 579
      ::::::::::::::::::::
512441 CACAAAAACAGATACAAATCCGAAATCCGACGTCGTGCAACCTTACTGTG 512490
      ::::::::::::::::::::
579 laAsnThrLeuTrpAsnThrTyTrpSerAspMetGlnAlaValGlsSerMet 595
      ::::::::::::::::::::
512491 CTACACACATATGGGATCCCTTGTGTGCGCTCCATCAACAGCTT 512540
      ::::::::::::::::::::
596 IleAsnThrIleAlaHisGlyAlaIleTyLeuPheGlyThrTrpGlySe 612
      ::::::::::::::::::::
512541 GTAGCCACTTAAGTACGCCAATCTCAGAAACTCGCGGACTGTGTGTA 512590
      ::::::::::::::::::::
612 rAlaValSerAsnLeuPheTyAlaHisAspSerSerGlySerProIleA 629
      ::::::::::::::::::::
512591 AGGATCTCGACACTTCTTC...CATAAAGATACGACGAAAGATTAATA 512634
      ::::::::::::::::::::
629 sPAsnTrpHisHisArgSerLeuGlyTyTrpLeuPheGlyIleSerThrHis 645
      ::::::::::::::::::::
512635 AAGGTTTCGCGCACATTAAGTGCAGGTTATGTGTAGAGCGCACATCAACA 512684
      ::::::::::::::::::::
646 SerLeuAspAspHisSerPheCysLeuAlaIleGlnLeuLeuGlyLy 662
      ::::::::::::::::::::
512685 TTAGCTTGTGAATCTTATATCATCTGCAGCTTCCTGCCAAATTAATCGGAA 512734
      ::::::::::::::::::::
662 sSerSerAspSerPheIleThrSerThrGluThrThrSerTyIleAlaI 679
      ::::::::::::::::::::
512735 AGAATAGACATCATTTAATAAATAAATAGACCTTCCTCCATATCAGCTT 512784
      ::::::::::::::::::::
679 hrValGlnAlaGln...LeuAlaThr.....SerLeuMetlys 690
      ::::::::::::::::::::
512785 CTTCATATCTCCACATCTACGACCTTGTCTCTCCAAAGCTTGTACGC 512834
      ::::::::::::::::::::
691 .....IleSerAlaGlnAl 695
      ::::::::::::::::::::
512835 TACCTTCTGATCTGAAGTACAGGCTTCCTCTTTGATGCTCAAT 512884
      ::::::::::::::::::::
695 acySTyAsnGluSerIleHisGluLeuSergIleTyTrpArgSerPheS 712
      ::::::::::::::::::::
512885 CAGCATATCTATAGTAAATAATAGATGAACCATATACACCAAGAC 512934
      ::::::::::::::::::::
712 eTyGluGlyPheGlySerTrpHisSerValAlaValSerGlyGluVal 728
      ::::::::::::::::::::
```

```

1259 GAGAGAGAGCTTCGAGACGACGAGAACTGCCGAAGCTGGATATATCTCAAAATCT 1308
      :::::::::::::::::::: || :::: :::::::::::::::::::: |||
415 GlnLeuGlnGlnProLeuLeuLysSerGlyArgValLeuLysAs 431
      :::::::::::::::::::: ||| :::::::::::::::::::: ||:::::
1309 ACAATTTCACCAACCTCTACCTCTTCGGAGGAGGCAACTCTCTTAATTC 1358
      :::::::::::::::::::: ||| :::::::::::::::::::: ||:::::
431 pArgValValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeu 448
      :::::::::::::::::::: ||| :::::::::::::::::::: ||:::::
1359 AGAGAGCATCTAGTTCGCAATCTCTTTCGCAATCTCCGGACTCTACC 1408
      AGAGAGCATCTAGTTCGCAATCTCTTTCGCAATCTCCGGACTCTACC
448 euLLeuMetLysAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeu 464
      :::::::::::::::::::: ||:::::::::::::::::: ||:::::
1409 TCCCTCAGGATCCAGGACCAATTAAGAACCCTGGATGGATCACTATC 1458
      TCCCTCAGGATCCAGGACCAATTAAGAACCCTGGATGGATCACTATC
465 AlaThrLeuSerLeuProLeuLysSerLeuAspThrGlnLysSerValThr 481
      :::::::::::::::::::: ||::: :::::::::::::::::::: ||:::::
1459 AATAACTCTCTTCTCAATCTAGATTCCTTAAAGAGACCAACAAAGCACTAC 1508
      :::::::::::::::::::: ||::: :::::::::::::::::::: ||:::::
481 rLLeuAlaAlaProAsnLeuSerLeuGlnLysLLeuLeuSerAsnSer 497
      :::::::::::::::::::: ||| ||| :::: :::::::::::::::::::: |||
1509 GCTAAAGCAACACAAAGCAAGT...CAGACAGCTACTTATCTGGATGAGC 1558
      GCTAAAGCAACACAAAGCAAGT...CAGACAGCTACTTATCTGGATGAGC
498 .....GlyAspGluAsnProPheGlyGlnAsnValGluLeuLeu 509
      :::::::::::::::::::: ||| :::::::::::::::::::: |||
1556 TCTCTCTTGTACATCTCTTTCGGAAATCTGTACGAAGATCTC..... 1596
      TCTCTCTTGTACATCTCTTTCGGAAATCTGTACGAAGATCTC.....
510 SerLysGluGlnAsnAsn.....LeuProLeuThrLeuSerLys 523
      SerLysGluGlnAsnAsn.....LeuProLeuThrLeuSerLys
1597 .....TCTTGCAGATTAACCTCTAGCTCTTTTCTGTCTCAGCTTTACTCTC 1640
      .....TCTTGCAGATTAACCTCTAGCTCTTTTCTGTCTCAGCTTTACTCTC
523 sGIuGln.....SerHisLeuHisLeuProAspGlyAsnLeuSerSer... 537
      :::::::::::::::::::: ||::: ||::: :::::::::::::::::::: |||
1641 TGACAGACCCCGGAATATTCACATCAACAAC.....TTAGCTGGTATC 1680
      TGACAGACCCCGGAATATTCACATCAACAAC.....TTAGCTGGTATC
538 .....HisPheGlyTyrGlnGlyAspThrPhe 547
      .....HisPheGlyTyrGlnGlyAspThrPhe
1685 CCCGTAGAAAAAATCCTATCCATTTGGGGATTCACCAAGGAAATGGGCATTGA 1734
      CCCGTAGAAAAAATCCTATCCATTTGGGGATTCACCAAGGAAATGGGCATTGA
548 SerTrpLysAspSerArgLysGlyHisSerLeuIleAla.....AsnTr 562
      SerTrpLysAspSerArgLysGlyHisSerLeuIleAla.....AsnTr
1735 TCTTGGCAAGAGATTACTCGCACTAATTCCAAAGCAGCCACTTTACTCTG 1784
      TCTTGGCAAGAGATTACTCGCACTAATTCCAAAGCAGCCACTTTACTCTG
562 rPThrProLysAsnTrpValProHisProGluArgGlnSerThrLeuValA 579
      rPThrProLysAsnTrpValProHisProGluArgGlnSerThrLeuValA
1785 GACAAAAACAGGATACAAATCCGATCTCAGCGGTGTGGAACTTAGTGT 1834
      GACAAAAACAGGATACAAATCCGATCTCAGCGGTGTGGAACTTAGTGT
579 laAsnThrLeuTrpAsnThrTrpSerAspMetGlnAlaValGlnSerMet 595
      laAsnThrLeuTrpAsnThrTrpSerAspMetGlnAlaValGlnSerMet
1835 CTACACACGCTATGGGATCTCTTGTGTGAGTGCCTCCCTACCAACAGCTT 1884
      CTACACACGCTATGGGATCTCTTGTGTGAGTGCCTCCCTACCAACAGCTT
596 IleAsnThrIleLeuHisGlyAlaIleArgLeuPheGlyThrTrpGlySe 612
      IleAsnThrIleLeuHisGlyAlaIleArgLeuPheGlyThrTrpGlySe
1885 GTAGCCACGTAAAGTATGCGCAATCTCAGAAACTCCGGGCACTTGGTGTGA 1934
      GTAGCCACGTAAAGTATGCGCAATCTCAGAAACTCCGGGCACTTGGTGTGA
612 rAlaValSerAsnLeuPheTrpValAlaHisAspSerSerGlyLysProIleA 629
      rAlaValSerAsnLeuPheTrpValAlaHisAspSerSerGlyLysProIleA
1935 AGGAGATCTCGAACTCTTC.....CATTAAGATAGCAGGAAGATAAATA 1976
      AGGAGATCTCGAACTCTTC.....CATTAAGATAGCAGGAAGATAAATA
629 sPAsnTrpHisHisArgSerLeuGlyTyrLeuPheGlyLysLeuThrHis 645
      sPAsnTrpHisHisArgSerLeuGlyTyrLeuPheGlyLysLeuThrHis
1979 AAGGTTTTCGCCACATTAAGTGCAGGTTTAGTGTAGGACCGCATCAACA 2022
      AAGGTTTTCGCCACATTAAGTGCAGGTTTAGTGTAGGACCGCATCAACA
646 SerLeuAspAspHisSerPheCysLeuAlaIleGlnLeuLeuGlyLys 662
      SerLeuAspAspHisSerPheCysLeuAlaIleGlnLeuLeuGlyLys
2029 TTAGCTTCTGATATCTTATTCACTGACGAGCTTCTGCCAATTATTCGGGAA 2076
      TTAGCTTCTGATATCTTATTCACTGACGAGCTTCTGCCAATTATTCGGGAA
662 sSerSerAspSerPheIleThrSerThrGlnThrHisSerTrpIleAlaTr 679
      sSerSerAspSerPheIleThrSerThrGlnThrHisSerTrpIleAlaTr
2079 AGATTAGAGATCACTTTAATAATTAATAATAGAGCTTCTGCCCTTAGCAGCT 2128
      AGATTAGAGATCACTTTAATAATTAATAATAGAGCTTCTGCCCTTAGCAGCT
679 hValGlnAlaGln...LeuAlaThr.....SerLeuMetLys 690
      hValGlnAlaGln...LeuAlaThr.....SerLeuMetLys
      :::::::::::::::::::: ||| :::::::::::::::::::: |||

```

159 HisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerLysGlyAsnGlu 175
1409 TCGAACTAT...AGTTGCTACTTGGCCAAACACTTTCTTAATGACAAATCG 600
175 YGlyAlaIleGlnAlaGlnThrPheSerLeuSerArgAsnValSerProI 192
601 AGCGCCCTCCAGGACGACTATGCTATGCTTAACCCCAAC...C 647
192 IeSerPheAlaArgAsnArgAlaAspLeuAsnGlyAlaIleLysGlyS 208
648 TAACCTTTGGCCAAAACAAAGCAGCAAGGAGGGGTGCTCTATTC 697
209 SerAsnLeuIleCysSerGlyAsn...ValAsnProLeuPhePheThrGlu 224
698 ACGGAGGAGATTACAATTAACTATGCTTAACATGATCATTTTCGA 747
224 yAsnSerAlaThrAsn... 229
748 AAATACCGCGGAGCAATGCGGAGCCATTACAGGAGAGCTACAGTT 797
229 229
798 TTATTAGCAGCAACAAAGCAATTAAGCTTTATTAACAAATAGTGACCCGA 847
230GlyGlyAlaIleCysCysIleSerAspLeuAsnThrSe 242
848 ACCTACGCTACAGGAGGAGCAATTTACTGCTACT...ACATGACAC 891
242 rGluLysGlySerLeuSerLeuAlaCysAsnGlnIleuThrLeuPheAla 259
892 CCCCAGGAGCTTAATCTATCATGACAGCAAGGAGGAGCACTTATATAG 941
259 erAsnSerAlaLysGluLysGlyAlaIleTyrAlaLysHisMetVal 275
942 GAAATACAGCAATTAAGTGGTGGGCAATTATATCTGACAACTAGTT 991
276 LeuArgTyrAsnGlyProValSerPheIleAsnAsnSer... 288
992 CTTTCTTCGAGGAGCACTACCTTTTAAACAACTCTGCGCTATGATAC 1041
289 ...AlaLysIleGlyGlyAlaIleAlaIleGlnSerGlyLysLeuS 304
1042 TGCAGCTCCCTTAGAGGAGGAGCAATGCGATCTGCTGATCTTGA 1091
304 erIleLeuAlaGluGluGlySerValLeuPheGlnAsnAsnSer... 318
1092 GTCTTCGCTCTTGGTGAGACATCACTTTTGAAGGAACACAGTAGTC 1141
319 ...GlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeu... 333
1142 AAAGGAGCTTCTCGAGTCAGACCACTACAGAAATTCATTAACATCGG 1191
334 .GluLysAspAlaIleLeuSerSerLeuGluAlaArgAsnGlyAsp...I 349
1192 AAACACCAATGCTAGATTTGTACAGCTGCGAGCTCTCAAGCAATACTA 1241
349 IeLeuPheAspProIleValGlnGluSerSerLysGluSerPro 365
1242 TCTACTTCTATGATCTATTAACA... 1264
366 LeuProSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAl 382
1265ACTAGCATCTGACGCTCTCTAGATGCTTAACCTAAATAG 1308
382 aserProLeuValIleGlnThrSerAlaAsnArgSerValIlePheSerS 399
1309 TCCCTGACCTTGCAGGGAATCTGCATATCAAGAACCAATGATTTTCG 1358
399 erGluArgLeuSerGluGluGluLysThr...ProAspAsnLeuThrSer 414
1359 GAGAGAGCTCTCGAAGCAGAGCTGCAAGAGCTGATATCTCAATCT 1408

415 GlnLeuGlnGlnProIleGlyLeuLysSerGlyArgLeuValLeuLysAs 431
1409 ACAATTCAGCAACTCTAACGCTTGGCGGAGGAGGCAACTCTCTTAATATC 1458
431 pArgAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeuL 448
1459 AGGAGTCACTTGTAGTCTTGTGCAATCTCCGGGCTCTACCC 1508
448 euIleMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeu 464
1509 TCTCTATGATGACAGGACCACTTGAACCGCTGATGGATCTACTATC 1558
465 AlaThrLeuSerIleProLeuHisSerLeuAspThrGluLysSerValTh 481
1559 AATAATCTCTTCATCATGATTCCTTAAGAGACCAAGAGGAGCAC 1608
481 rIleHisAlaProAsnLeuSerIleGlnLysIlePheLeuSerAsnSer 497
1609 GCTAAAGCAACACAGCAAGT...CAGACAGTCACTTATCTGATGATCGC 1655
498GlyAspLysAsnPheGlyGluAsnValGluLeuLeu 509
1656 TCTCTCTTGTAGATCTCTGGAATGCTTAGAGATGTC... 1696
510 SerLysGluGlnAsnAsn...IleProLeuLeuThrLeuSerLys 523
1697TCTGGAAATACCCCTCAAGTCTTCTGTGCTCACTCTTACTGC 1740
523 sGluGln...SerHisLeuHisLeuProAspGlyAsnLeuSerSer... 537
1741 TGAGGACCCCGCAATATATCATCACAGAC...TTACTCTCTGATC 1784
538HisPheGlyTyrGlnGlyAspTyrPhe 547
1785 CCCCAGAAAAATCTCATCTGATGGGATACCAAGGAAATGGGCAATTA 1834
548 SerTrpLysAspSerAspGluGlyHisSerLeuIleAla...AsnTr 562
1835 TCTTGCAAGAGATATCGCGCAATATCCAAAGAGGAGGACTTACTCTG 1884
562 pThrProLysAsnTyrValProHisProGluArgGlnSerThrLeuVala 579
1885 GACAAAAACAGATATCAATCCGAATCCGAGCGTCGTGGAACCTTAGTTG 1934
579 IaAsnThrLeuTrpAsnThrTyrSerAspMetGlnAlaValGlnSerMet 595
1935 CTACACCCCTATGGGATCTTGTGATGTGCGCTCATACAAACAGCTT 1984
596 ILeAsnThrIleAlaHisGlyAlaTyrLeuPheGlyThrTrpGlySe 612
1985 GTACCCACTAAAGTACGCCAATCTCAAGAACTCGGCGCATCTGGTGA 2034
612 rAlaValSerAsnLeuPheTyrAlaHisAspSerSerGlyLysProIleA 629
2035 AGGGATCTCGAACTTCTC...CATAAAGATAGCAGGAATTAATA 2078
629 sPasnTrpHisThrArgSerLeuGlyTyrLeuPheGlyIleSerThrHis 645
2079 AAGCTTTTCGCCACATAAGTCAGGTTATGTTGTAGGAGGCACTACACA 2128
646 SerLeuAspHisSerPheCysLeuAlaIleGlnGlnLeuLeuGlyLy 662
2129 TTAGCTTCTGAATATCTATACCTGACGCTTTCGCCAATTAATTCGGCA 2178
662 sSerSerAspSerPheIleThrSerThrGluThrThrSerTyrIleAlaT 679
2179 AATAGATGATCATCTTATTAATAAAGAGCTTCTCCATAGCAGCTT 2228
679 hrValGlnAlaGln...LeuAlaThr...SerLeuMetLys 690
2229 CTCTCATCTCCAGCATCTAGGAGCTTGTCTTCCAGAGCTTGTACGC 2278
691IleSerAlaGlnAl 695

192 leSerPheAlaIArgAsnARgAlaAspLeuAnsglyAlaIleCysCys 208
1409 TACGCTTTGCCAAACAAACAGCAGCCAAAGGCGCGCCCTCATTTCC 597
209 SerAsnLeuIleCysSerGlyAsn...ValAsnProLeuPhePheThrG1 224
598 ACGGAGGAGATTACAAATTAAACATACGTTAAACATCATTTCTGGA 647
224 yAsnSerAlaThrAsn..... 229
648 AAATACCGCGGGAACAATGGCGGACATTACACGGAACCTACGACTT 697
229 229
698 TTATTACGACCAACAACAAATTAGCTTTATTAACAATAGTGCACGCCA 747
230GlyGlyAlaIleCysCysIleSerAspLeuAsnThrSe 242
748 ACCTGAGCTACAGGGGGAGCCATTACTAGTGT.....ACATCAGC 791
242 rGluIleGlySerLeuSerIleuAlaCysAsnGlnGluThrLeuPheAla 259
792 CCCCAGAACCACTTAACTCTATACGACAGCGGGAACCTGAACTTTATG 841
259 erAsnSerAlaLysGluLysGlyAlaIleTyAlaLysIleMetVal 275
842 CAATATACCAAAATTACTAGTGGCGCCCATTTATACGACAACTCTGAT 891
276 LeuArgTyIraAsnGlyProValSerPheIleAsnAsnSer..... 288
892 CTTCTCTGAGAGACCTACGCTTTTAAACAACTCTGCTATGATGATC 941
289 ...AlaLysIleGlyAlaIleAlaIleGlnSerGlyGlySerLeu 304
942 TGCAGCTCCTTAGAGAGCAATGGCATTTGCTGACTCTGATCTTGA 991
304 erIleLeuAlaGlyGlnGlySerValLeuPheGlnAsnAsnSer..... 318
992 CTCTTTCGCTCTGGTGGAGACATCACTTTTGAAGCAACACAGTACTC 1041
319 ...GlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyLeu.. 333
1042 AAAGAGCTTTCGAGTCAAGCACTACAGAAATTTCTATTACATCGG 1091
334 .GluLysAspAlaIleLeuSerSerLeuGluAlaArgAsnGlyAsp... 349
1092 AAACACCAATGCTAAGATTGTACACAGCTCGAGCCCTCAAGCATTACTA 1141
349 leLeuPhePheAspProIleValGlnIleSerSerLysGlnSerPro 365
1142 TCTACTCTATGATCTTATACA..... 1164
366 LeuProSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAl 382
1165ACTAGATCATCTGCAGAGCTCTCAAGATGCTTAACCTTAAGG 1208
382 aserProLeuValIleGlnThrSerAlaAsnArgSerValIlePheSer 399
1209 TCCTACTCTTGAGGAGATCTCTGATATCAAGAACCATGATTTCTG 1258
399 erGluArgLeuSerGlnGlnGlyLysThr...ProAspAsnLeuThrSer 414
1259 GAGAGAACTCTCGAAGACAGACCTCCAGAACTGATAATCTCAATCT 1308
415 GlnLeuGlnGlnProIleGlnLeuLysSerGlyArgLeuValLeuLys 431
1309 ACAATTACGCAACTCTACTCTGCGGAGGAGCACTCTCTTAATATC 1358
431 parGAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLul 448
1359 AGGATCTACTCTAGTTGCTAAGTCTTTTCGAAATCTCCGGGCTTACCC 1408

448 euIleMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeu 464
1409 TCCTCATGATCGACGAGGACACATTAAGAACCCCTGATGCATCATATC 1458
465 AlaThrLeuSerIleProLeuHisSerLeuAspThrLysSerValTh 481
1459 AATATCTTCTTCAATGTAATTCCTTAAGACCAAGCAAGGCGAC 1508
481 rIleHisAlaProAsnLeuSerIleGlnLysIlePheLeuSerAsnSer 497
1509 GCTAAAGCAACACACCAACACT...CAGACAGTCACCTTTATCTGATGCC 1555
498GlyAspLysAsnPheTyGlnAsnValGluLeuLeu 509
1556 TCCTCTCTGATAGCTCTTCAAGATCTCAAGATCTC..... 1596
510 SerLysGlnGlnAsnAsn.....IleProLeuLeuThrLeuSerLy 523
1597TCTTGAAATACCCCTCAAGCTTTCTTGTCTACTCTTACTGC 1640
523 sGlnGln...SerHisLeuHisLeuProAspGlyAsnLeuSerSer.... 537
1641 TGACAGACCCCGCAATATTCATCATCACAGAC.....TTAGCTCTGATC 1684
538HisPheGlyTyGlnGlyAspTyrThrPhe 547
1685 CCCYAGAAAAAATCTATCCATGCGGATACCAAGCAATGGGCAWTA 1734
548 SerTyrLysAspSerAspGlnGlyHisSerLeuIleAla.....AsnTyr 562
1735 TCTTGGCAAGGAGTACTGCGACTAAATCCAAGACAGCACTTACTCTG 1784
562 pThrProLysAsnTyAlaProHisProGluArgGlnSerThrLeuVal 579
1785 CACAAACAGACATACAAATCCGAAATCTCGAOCCTGCGAACCCTTACTG 1834
579 laAsnThrLeuTyrAsnThrTySerAspMetGlnAlaValGlnSerMet 595
1835 CTACACCGCTATCGGATCTCTTGTGATGCGCCTCAATACACACACTT 1884
596 IleAsnThrIleAlaHisGlyAlaIleTyLeuPheGlyThrTyrGlyse 612
1885 GTACCCCTAAAGTAAGCCCAATTCAGAAACCTCGCGCACTTGTGTGA 1934
612 rAlaValSerAsnLeuPheTyAlaHisAspSerSerGlyLysProIleA 629
1935 AGGATCTCGAATCTCTTC.....CATAAAGATAGACAGCAAGTAATA 1978
629 sPAsnTyrHisHisArgSerLeuGlyTyLeuPheGlyIleSerThrHis 645
1979 AAGCTTTCCGCCACATTAAGTGCAGCTTATGTTGTAGAGAGCACTAACA 2028
646 SerLeuAspAspHisSerPheCysLeuAlaIleGlnLeuLeuGlyLy 662
2029 TTAGCTCTGATATCTTATCTACGACAGCTTTCGCAATTAATTCGGGA 2078
662 sSerSerAspSerPheIleThrSerThrGlyThrThrSerTyIleAla 679
2079 AGATAGAGATCACTTAAATAAATAAATAGAGCTTCGCCATATGCACCT 2128
679 hrValGlnAlaGln...LeuAlaThr.....SerLeuMetLys 690
2129 CTCTTCATCTCCACATCTAGCAGCTTGTCTCTCAAGCTTGTTCAGC 2178
691IleSerAlaGlnAl 695
2179 TACCTCTCGATCTGAAGAGAGAGGCTCTCTTGTATGCTCGAGAT 2228
695 acTyTyraAsnGlnSerIleHisGlnLeuLysThrLysTyraSerPheS 712
2229 CACGATATCTATAGTAAATAATAGTAAGAACTATAGACCAAGAC 2278
712 erLysGlnGlyPheGlySerThrPheSerValAlaValSerGlyLulVal 728

105091 AATATACCGCGGCAACAATGGCGGAGCCATTACAGGAGACCTAGCACTT 105042
 229
 105041 TTATTAGCAGCAACAAACATTAAGCTTTATTAACATAGTGTGACCCCA 104992
 230glyAlaIleCysCysIleSerAspLeuAsnThrSe 242
 104991 ACCTCAGCTACAGGGGAGGACATTAAGTACTACT.....ACATCAGC 104948
 242 rGluLysIleSerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAla 259
 104947 CCCCACAAACCTTAACTTAAGTACAGACAGCGGAACTGAATGAACTTAAAG 104898
 259 eAsnSerAlaLysGluLysGlnGlyAlaIleTyrAlaLysIleMetVal 275
 104897 GAATACAGCAATTAAGTACTAGTGTGGGCGGATTTATAGTACATTAAGTT 104848
 276 LeuAlaGlyAsnGlyProValSerPheIleAsnAsnSer..... 288
 104847 CTTCCTTCGAGAGCACTACCGCTTTTAAACAACTCTCGATAGATAC 104798
 289 ...AlaLysIleGlyAlaIleAlaIleGlnSerGlySerLeu 304
 104797 TGCAGCTCCCTAGAGAGCAATGCGATGCTGACCTGATCTTGA 104748
 304 eTlleLeuAlaGlyGlyLysSerValLeuPheGlnAsnAsnSer..... 318
 104747 GTCTTGGCGCTCTGGTGAGACATCACTTTGAAAGAAACACATAGTC 104698
 319 ...GlnAlaGlyThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeu 333
 104697 AAAGACCTCTTCGAGTACAGACCACTACAGAAATCTTAATTAACATCGG 104648
 334 .GluLysAspAlaIleLeuSerSerLeuGluAlaArgAsnGlyAsp...I 349
 104647 AAACCCCAATCTAAGATTGATGACAGCTCGAGCCCTCAAGCACTACTA 104598
 349 lLeuPhePheAspProIleValGlnGlnSerSerGlySerPro 365
 104597 TCTACTTCTATGATCTTATACA..... 104575
 366 LeuProSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAl 382
 104574ACTACATCAGTCGAGACGCTGCAGATGCTAAACTTAATAGC 104531
 382 aSerProLeuValIleGlnThrSerAlaAsnArgSerValIlePheSer 399
 104530 TCCTGACCTTCAGGGAATCCTGCTATCAAGAACATCGATTTTCTG 104481
 399 eArgLysLeuSerGlnGluGluLysThr...ProAspAsnLeuThrSer 414
 104480 GAGAGACCTCTCGAGACAGAACTGCAGAACTGATTAATCTCAATCT 104431
 415 GlnLeuGlnProIleGluLeuLysSerGlyArgLeuValLeuLysAs 431
 104430 ACAATTCAAGCAACCTTAACCTTCGCGAGGCAACTCTCTTAATATC 104381
 431 pArgAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeu 448
 104380 AGAGCTCACTAGTCTAGTCTTTCCTTCCAACTCCGGGCTCTTACC 104331
 448 eLleMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeu 464
 104330 TCCTATGATGATGAGGAGACCATTAAGAACCGCTGATGGATCTCAATC 104281
 465 AlaThrLeuSerIleProLeuHisSerLeuAspThrGlyLysSerValTh 481
 104280 AATAATCTTGTTCATCTAGATCTTAAAGAGACCAAGAACAGCTAC 104231
 481 rLleHisAlaProAsnLeuSerIleGlnLysIlePheLeuSerAsnSer 497
 104230 GCTAAAGCAACACAGCAAGT...CAGACAGTCACTTATCTGGATCG 104184
 498GlyAspGlnAsnPheTyrGluAsnValGluLeu 509
 104183 TCTCTCTGTAGATCCCTTCTGAAATGCTCTACGAAGATCTC..... 104143
 510 SerLysGlnGlnAsnAsn.....IleProLeuLeuThrLeuSerLys 523
 104142CTTCGAATAACCCCTCAAGTCTTCTGTGTCTACTCTTACTTC 104099
 523 sGluGln...SerHisLeuHisLeuProAspGlyAsnLeuSerSer... 537
 104098 TGACGACCCCGGAATATTACATCAGAC.....TTCACCTCTGATC 104055
 538HisPheGlyTyrGlnGlyAspThrPhe 547
 104054 CCGTAGAAAAAATCCATTCATTCGGGATACCAAGGAATGGGCACTA 104005
 548 SerTrpLysAspSerAspGlnGlyHisSerLeuIleAla.....AsnTr 562
 104004 TCTTCGCAAGAGATACGACGACATAATCCAAACAGCACTCTTACCTG 103955
 562 pThrProLysAsnTyrValProHisProGluArgGlnSerThrLeuVal 579
 103954 GACAAAAACAGATACATCCGAATCTGAGCGTGTGAACCTTAGTGG 103905
 579 lAsnThrLeuTrpAsnThrTyrSerAspMetGlnAlaValGlnSerMet 595
 103904 CTACACCCCTTGGGAGATCCTTGTGTGATGTGCGCTCATCAACAGCTT 103855
 596 lLeasnThrIleAlaHisGlyAlaTyrLeuPheGlyThrTrpGlySe 612
 103854 CTAGCCACTAAGTACGCCAATCCAAAGAACTCGCGCATCTGTGCGA 103805
 612 rAlaValSerAsnLeuPheTyrAlaHisAspSerSerGlyLysProIleA 629
 103804 AGGGATCTCGAATCTTCTC.....CATTAAGATAGCAGGAATATA 103761
 629 sPasnTrpHisIleArgSerLeuGlyTyrLeuPheGlyIleSerThrHis 645
 103760 AAGCTTTGCGCACATTAAGTGCAGCTTATCTTTAGAGACCGACATCAACA 103711
 646 SerLeuAspAspHisSerPheCysLeuAlaAlaGlyLeuLeuGlyLys 662
 103710 TTAACCTCTGAAATCTTATCACTGCACCTCTGCTCCAAATTAATTCGCGAA 103661
 662 sSerSerAspSerPheIleThrSerThrGluThrSerTyrIleLeuAla 679
 103660 AGATAGAGATCACTTATTAATAATAAATAGAGCTTCTGCTATGACAGCTT 103611
 679 hrValGlnAlaGln...LeuAlaThr.....SerLeuMetLys 690
 103610 CTCTCCATCTCCAGCATCTAGCAGACCTTGTCTCTCCAGCTTGTACGC 103561
 691IleSerAlaGlnAl 695
 103560 TACCTCTCGATCTGAAGAGTGCAGAGCCCTGCTCTTGTATGCTCAAGAT 103511
 695 acsTyrTrpAsnGlnSerIleHisGluLeuLysThrLysTyrArgPheSer 712
 103510 CAGCTATATCTATAGTAAATACTATGAAACCTATTACACCCCAAGCAC 103461
 712 eLysGluGlyPheGlySerTrpHisSerValAlaValSerGlyGluVal 728
 103460 CAAC...CGAGAGAGCTCGTGTATATAGCGGTGGCGCTCGAAGCTT 103414
 729 CysAlaSerIleProIleValSer...AsnGlySerGlyLeuPheSerSe 744
 103413 GCGAGCTCCCTACACACACATGCTTAAGCCATGAGCGGCTCTTCTCCAGC 103364
 744 rPheSerIlePheSerLysLeuGlnGlyPheSerGlyThrGlnAspGlyP 761
 103363 GTATTTCTTTCATCAAGTAGAGCTTCTGATACATACCAAGATAGCT 103314

[illegible]

1228	CTGAAATCCGAGCGCTTACTTTTAAAGATCGCGCTGTCCTTTCCGSGCC	1277
438	oSeRLeuSerGlnAspProGlnAlaLeuLeuLleMetGlnAlaGlyThrs	455
1278	TTCTGCTCTCGAGATCCCAAGGCTCTCTCTATTGGAAGCGGAACTT	1327
455	eRLeuLysThrSerSerAspLeuLysLeuAlaThrLeuSerLleProLeu	471
1328	CTTAAAAACTCTCTGTGATTGAAGTTAGSTAGACGTAAGTATCCCTT	1377
472	HisSerLeuAspThrGluLysSerValThrLleHisAlaProAsnLeuSe	488
1378	CAATCCCTTGAAATACGAAAAAGCGCTACTATCCACGCCCTAACTTTC	1427
488	rLleGlnLysLlePheLeuSerAsnSerGlyAspLysAsnPheTyrgLva	505
1428	TATCCAAAAGATCTTCTCTCTCTAAGCTTGGAGATAGAAATTTTATGAA	1477
505	sNValGluLeuLeuSerLysGluGlnAsnAsnLleProLeuLeuThrLeu	521
1478	ATGTAGAGCTTCTCAGTAAAGAGCAAAACAATATTCCTCTCTTAAGTCC	1527
522	SeLysGluGlnSerHisLleuLisLeuProAspGlyAsnLeuSerSerH	538
1528	CCTAAAGACCAATCTCAATTAACATCTTCTGTATGGGAACCTTCCTTCA	1577
538	sPheGlyTyrgInGlyAspTrpCThrPheSerTrpLysAspSerAspGlu	555
1578	CTTTGGATATCAAGAGAGATTGGACTTTTCTTGGAAAGATTCTGATGAG	1627
555	LHisSerLeuLleAlaAsnTrpThrProLysAsnTyrgValProHisPro	571
1628	GGGATTCCTGATTCGATTTGGAGCGCTAAAAACTATGTGCTATGCA	1677
572	GluArgGlnSerThrLeuValAlaAsnThrLeuTrpAsnThrTyrgSerAs	588
1678	GAACTGCAATCTACACTCTCTCCGAACACTCTTTCGGAACACCTATTCGGA	1727
588	pMetGlnAlaValGlnSerMetLleAsnThrLleAlaHisGlyGlyAlaT	605
1728	TATGCAAGCTGTGCAGTGCATATTAATACACAGCAGCAGGAGGAGCTT	1777
605	yrLeuPheGlyThrTrpGlySerAlaValSerAsnLeuPheTyrgLHis	621
1778	ATCTATTGGAACGAGGGGATGTGCTGTTCTTAATTATTCATGTCTGAC	1827
622	AspSerSerGlyLysProLleAspAsnTrpHisHisArgSerLeuGlyTy	638
1828	GACAGCTCTGGGAAACCTATCCATTAATTGGCATTAAGAGCCTTGCGCTA	1877
638	rLeuPheGlyLleSerThrHisSerLeuAspHisSerPheCysLeuA	655
1878	CCATTCGGTATACAGTACCAAGATTGAATACCAATCTTCTGCTTGG	1927
655	lAlaAlaGlnLeuLeuGlyLysSerSerAspSerPheLleHisSerThr	671
1928	CTGCAGACCAATTACTCGGGAAATGTCGCGATTCTTATTACAGCTACA	1977
672	GluThrThrSerTyrlAlaAlaThrValGlnAlaGlnLeuAlaThrSerLe	688
1978	GAACACACCTCCCAATATACTACTGTACAGCGCAACCTGCACTCTCT	2027
688	uMetLysLleSerAlaGlnAlaCysTyrgAsnGluSerLleHisGluLeuL	705
2028	ATGTAAAACTCTGCAAGCAAGCATGCTACATGAAGTATCCATGTGCTTAA	2077
705	ysThrLysTyrgArgSerPheSerLysGluGlyPheGlySerThrPheSer	721
2078	AAACAATAATATGCTCTCTCTTAAGAAGAGATTGGATCTCGCATATGC	2127
722	ValAlaValSerGlyGluValCysAlaSerLleProLleValSerAsnGly	738
2128	GTTCGCGATTCGCGGAAGATGTGGCCATCGATGATCTTATGTATCCAAATGG	2177

472013 ATGAGAGCCTCTTCTCTCTTGTGCTCATGATCACAAGACGATCC 472062
67 roleuTyRValleuGlyAsnSerTyCysTrpPheValSerLysLeuHis 83
|||||
472063 CTTGTATGACTGGAATTCCTACTGTTGCTGATCTTAACCTCAT 472112
84 lIethAsPProIySGuAlaLeuPheLysGluLysGlyAspLeuSerl 100
|||||
472113 ATACAGACCCCAAGAGCGCTCTTTTAAAGAAAAAGAGATCTTCCAT 472162
100 eGIAsnPheArGpHeuSerPheThrAspCysSerSerLysGluSers 117
|||||
472163 TCAAACCTTCGCTCTCTTCTCTTCCACAGATGCTCTTCCAAAGAAACCT 472212
117 erProSerlIeIleHISGlnLysAsnGlyGlnLeuSerLeuArGAsnAsn 133
|||||
472213 CTCCTCTATATTATTCATCAAAAGATGTCAGTTATCTTGCCCAATAT 472262
134 GlySerMetSerPheCysSarGAsnHISAlaGluClySerClyAlaIle 150
|||||
472263 GGTAGCATGAGTTCTGTGCAATCATGCGAAGCGCTTGAGAGACCAT 472312
150 eSerAlaAsPAlaPheSerLeuGlnHISAsnTyRLeuPheThrAlaPheG 167
|||||
472313 CTCTGCGCATCCTTTTCTCTCTACAGCACAATCTTTTTCACACCTTTTC 472362
167 lUgLuAsnSerSerLysGlyAsnGlyAlaIleGlnAlaGlnThrPhe 183
|||||
472363 AAGAGATTCCTCTTAAGGAAATGGCGAGCCATTCAGGCTCAACCTTC 472412
184 SerLeuSerArGAsnValSerProIleSerPheAlaArGAsnArGAlaAs 200
|||||
472413 TCTTTATCTAGAAATGTGCGCTATTCTTTCGCCCTAATGCTGGCGA 472462
200 pIeAsnGlyAlaIleCysCysSerAsnLeuIleCysSerGlyAsnTy 217
|||||
472463 TTTAATTCGCGCGCGCTATTTCCTGCTAGTAATCTTATTTGTCAGGAAATG 472512
217 aIAsnPProIeuPhePheThrGlyAsnSerAlaThrasnGlyAlaIle 233
|||||
472513 TAAACCTCTCTTTTCTACGCAACTCCGCCAGAAATGAGCGCTATT 472562
234 CysCysIleSerAspLeuAsnTrpSerGlyLysGlySerLeuSerLeuAl 250
|||||
472563 TGTGTATAGCGCATCTAAACACCTCGAAAAAGCGCTCTCTCTCTGCG 472612
250 acYAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysGluLysGlyG 267
|||||
472613 TTCTAACCAAGAACCGCTATTTCGAAGCAATTCGCTAAAGAAAAAGCGC 472662
267 lYAlaIleTyRAlaLysHISMetValLeuArGTyRAsnGlyProValSer 283
|||||
472663 GGGCTATTATGCGCAAGCACATGTTGCGTTATTAACGTCCTGTTTGC 472712
284 pHeIleAsnAsnSerAlaLysIleGlyAlaIleAlaIleGlnSerC 300
|||||
472713 TTCATTAAACAAAGCGCTAAATAGGTGAGCTATCGCATCCAGCCGG 472762
300 yGlySerLeuSerIleLeuAlaGlyGluLysSerValLeuPheGlnAsnA 317
|||||
472763 ACGGACTCTCTATCTTCAGGTGAGGATGCTGTTGTTCCAAAGAA 472812
317 snSerGlnArGThrSerAspGlnGlyLeuValArGAsnAlaIleTyRLeu 333
|||||
472813 ACTCCCAAGCAAGCTCCAGCAAGGCTGTAAGAAAGCCCATCTACTTA 472862
334 GluLysAsPAlaIleLeuSerSerLeuGlnAlaArGAsnGlyAspIleLe 350
|||||
472863 GAGAAAGATGACATTCTTCTCTTAGAAGCTCGCAAGGAGATATCT 472912
350 uPhePheAsPProIleValGlnGluSerSerSerLysGlySerProLeuP 367
|||||
472913 TTTCTTTGATCTCTATTGTACAAGAAAGTAGCAGCAAGAAATCCCTCTTC 472962

367 roSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAlaSer 383
|||||
472963 CTCCTCTTTTGCAAGCAGGTACTTCTCCACCCAGCACCGCATCT 473012
384 ProLeuValIleGlnThrSerAlaAsnArGSerValIlePheSerSerG 400
|||||
473013 CCTTGTATTATTCAGCAAGTCGCAAAACCGTTCAGTATTTTCTCGAGCA 473062
400 uArGLeuSerGluGluLysThrProAspAsnLeuThrSerGlnLeuG 417
|||||
473063 ACCTCTTTCGAGAGCAAAAAAACCCTGTATACCTGACTCCCACTAC 473112
417 lInGlnProIleGlnLeuLysSerGlyArGLeuValLeuLysAspArGAla 433
|||||
473113 AGCAGCTTATCGACTGAATCCGAGCGCTGTAGTTTAAAGATCGCGCT 473162
434 ValLeuSerAlaProSerLeuSerGlnAsPProGlnAlaLeuLeuIleMe 450
|||||
473163 GTCCTTCCCGCGCTCTCTCTCTCAGATCCCAAGCTCTCTCTATTAT 473212
450 lGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAlaThrL 467
|||||
473213 CGAACCGGAGACTTCTTAAAAACTCTCTGATTTGAAGTTACT ACCC 473261
467 euSerlIeProLeuHISerLeuAsPThrGluLysSerValThrIleHIS 483
|||||
473262 TAAGTTTCCCTTCACTTCTTAGATACTGAAAAAGCGTACTATCCAC 473311
484 AlaProAsnLeuSerlIeGlnLysIlePheLeuSerAsnSerGlyAspG 500
|||||
473312 GCCCTTAATCTTTCATCCAAAAAGATCTTCTCTCAACTCGAGATGA 473361
500 uAsnPheTyRGlusAsnValGluLeuLeuSerLysGluGlnAsnAsnIleP 517
|||||
473362 GAATTTTATGAAAAATGTAGACCTTCTCACTAAGAGCAAAACAATATTC 473411
517 roleuLeuThrLeuSerLysGluGlnSerHISLeuHISLeuProAspGly 533
|||||
473412 CTCTCTTACTCTCTCTTAAGAGCAATTCATTTACATTTCTCTGATGG 473461
534 AsnLeuSerSerHISPheGlyTyRGlusGlyAsPTrpThrPheSerTyRly 550
|||||
473462 AACCTCTTCTTACACTTGTGATATCAAGAGATGTGACTTTTCTTGGA 473511
550 sAsPSerAspGluGlyHISerLeuIleAlaAsnTrpThrProLysAsnT 567
|||||
473512 AGATTCTGAWGAAGGCAATTCCTGATTCGTAATTGGAAGCGCTAAAAACT 473561
567 TyRValProHISProGluArGlnSerThrLeuValAlaAsnThrLeuTrp 583
|||||
473562 ATGTGCTTATTCAGACGTCATCTACACTCGTTGCGAACCTCTTGG 473611
584 AsnThrTyRSerAspMetGlnAlaValGlnSerMetIleAsnThrIleAl 600
|||||
473612 AACACCTATTTCGATATGCAAGCTGTGAGTGTGATTAATACAAACAGC 473661
600 aHISGlyAlaIleTyRLeuPheGlyThrTrpGlySerAlaValSerAsnL 617
|||||
473662 GCACGAGAGACCGCTATCTATTGGAACGGAACGATCGCTGTTCTAAT 473711
617 euPheTyRAlaHISAspSerSerGlyLysProIleAspAsnTrpHIS 633
|||||
473712 TATCTATGTTTACGACAGCTCTGGAAAACTATGATTAATTTGGCATAT 473761
634 ArGSerLeuGlyTyRLeuPheGlyIleSerThrHISerLeuAspAspH 650
|||||
473762 ACAAGCCTTGCTACTATTCGCTATCACTACTCACTTATAGATGACCA 473811
650 sSerPheCysLeuAlaIleGlyGlnLeuLeuGlyLysSerSerAspSerP 667
|||||
473812 TTTCTTTGCTTGCTGCGAGCAATTTACTCGGAAATTCGTCCTTCTCT 473861

Align seg 1/1 to: A64751 from: 1 to: 2643

```
1 MetArgProAspHisMetAsnPhenCysCysLeuGlyAlaAlaIleLeuSe 17
|||||
7 ATGGAGCTGATCATATGACTTGTGTCTATGTCCTGCTATTTGTC 56
|||||
17 rSerThrAlaValLeuPheGlyGlnAspProLeuGlyGluThrAlaLeuL 34
|||||
57 ATCCACAGCGGCTCCTTTGGCCAGGATCCCTTAGTGAAACCGGCCCTCC 106
|||||
34 eutThrLysAsnProAsnHisValValCysThrPhePheGluAspCysThr 50
|||||
107 TCACATAAAATCTTAATCATGCTCTCTGTACATTTTGTAGACACTGTAC 156
|||||
51 MetGluSerLeuPheProAlaLeuGlyAlaHisAlaSerGlnAspAspR 67
|||||
157 ATGGAGAGCTCTTCTCCCTCTTGTCTCATGATCATCAAGAGAGATCC 206
|||||
67 oLeuTyValLeuGlyAsnSerTyGlyStrPheValSerLysLeuHisI 84
|||||
207 TTTGTATGACTTGGAATTCCTACTGTTGGTTCGTATCTAAACTCCATA 256
|||||
84 leThrAspProLysGluAlaLeuPheLysGlyLysGlyAspLeuSerIle 100
|||||
257 TCACGAGACCCCAAGAGGCTCTTTTAAAGAAAAGAGATCTTTTCCATT 306
|||||
101 GlnAsnPheArgPheLeuSerPheThrAspCysSerSerLysGlySerSe 117
|||||
307 CAAAACTTCCGCTTCTCTTCCCTCAGATGCTCTTCCAAAGAAAGCTC 356
|||||
117 rProSerIleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsnG 134
|||||
357 TCCTTCTATATTCATCAAAAGAAATGGTCAGTATTCCTTGGCCATATAG 406
|||||
134 LysSerMetSerPheCysArgAsnHisAlaGlyLysGlyValAlaIle 150
|||||
407 GTACCATAGTTTCTGCGAAATCATGCTGAAGGCTCTGAGACGCCATC 456
|||||
151 SerAlaAspAlaPheSerLeuGlnHisAsnTyLLeuPheThrAlaPheG 167
|||||
457 TCTGGGATGCTTCTTCTCTACAGACAACTATCTTTACAGCTTTTGA 506
|||||
167 uGluAsnSerSerLysGlyAsnGlyValAlaIleGlnAlaClnThrPheS 184
|||||
507 AGAGATTTCTTCTAAAGAAATGGCGGAGCCATTCAGGCTCAAAACCTTCT 556
|||||
184 eLeuSerArgAsnValSerProIleSerPheAlaArgAsnArgAlaAsp 200
|||||
557 CTTTATCTAGAAATGTCGCCCTATTTCTTCCGCCGTAATCTCCGGAT 606
|||||
201 LeuAsnGlyAlaAlaIleCysCysSerAsnLeuIleCysSerGlyAsnTy 217
|||||
607 TTTAAATGGCGGCGCTATTTGCTGTAGTAACTTATTGTTCAAGGAAATGT 656
|||||
217 AsnProLeuPhePheThrGlyAsnSerAlaThrAsnGlyAlaAlaIleC 234
|||||
657 AAACCCCTCTCTTTCTACTGGAACCTCCGCCAGAAATGAGGCGCTATTT 706
|||||
234 ySCysIleSerAspLeuAsnThrSerGlyLysGlySerLeuSerLeuAla 250
|||||
707 GTTGATACGGCATCTAAACACCTCAGAAAAGGCTCTCTCTCTGCT 756
|||||
251 CysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysGlyLysGly 267
|||||
757 TGTAAACCAAGAAACGCTATTTCAGACAACTCTGCTAAAGAAAAGGCGG 806
|||||
267 ValAlaIleTyAlaLysHisMetValLeuArgTyAsnGlyProValSerP 284
|||||
807 GCGCTATTATGCCAAGACATGTAATGCTTAATACGCTCTCTGTTTCT 856
|||||
284 hisLeuAsnSerAlaLysIleGlyAlaIleAlaIleAlaIleGlnSerGly 300
|||||
857 TCATTTAACACAGCGCTAAATAAGTGGAGCTATGCCCATCCAGTCCGGA 906
|||||
301 GlySerLeuSerIleLeuAlaGlyGlyGlySerValLeuPheGlnAsnAs 317
|||||
907 GGGAGTCTCTCTATCTCTTGCAGGTGAAGATCTGTTCTTCCAGAAATA 956
|||||
317 nSerClnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyLeuG 334
|||||
957 CTCCCAAGCAGCTCCGACCAAGGCTAGTAAGAAACCCCAATCACTTAR 1006
|||||
334 LuLysAspAlaIleLeuSerSerLeuGluAlaArgAsnGlyAspIleLeu 350
|||||
1007 AGAAAGATCCGATTTCTTCTCTTAGAAGCTCCCAAGCGAGATATCTT 1056
|||||
351 PhePheAspProIleValGlnGlySerSerSerLysGlySerProLeuP 367
|||||
1057 TTCTTTGATCTTATGTACAGAAAGTAGCAGAAAGATGCGCTTCC 1106
|||||
367 oSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAlaSerP 384
|||||
1107 CTCCCTTTGCAAGCCAGCGTACTCTCCACCCACCCACCGCATCTC 1156
|||||
384 rLeuValIleGlnThrSerAlaAsnArgSerValIlePheSerSerGlu 400
|||||
1157 CTTTAGTTATTCACACAGCTCAACGCTTCAGTCAATTTTCTCGACGCA 1206
|||||
401 ArgLeuSerGluGlyGlyLysThrProAspAsnLeuThrSerGlnLeuG 417
|||||
1207 GGTCTTTCTGAAGAAAGAAACCTCTGATTAACCTCACTTCCAACTACA 1256
|||||
417 nGlnProIleGlnLeuLysSerGlyArgLeuValLeuLysAspArgAla 434
|||||
1257 GCAGCCTATTCAGCAAGAAATCCGAGCGTTAGTTTAAATAATCCCGCTG 1306
|||||
434 allLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeuLeuIleMet 450
|||||
1307 TCCTTCCGCGCTCTCTCTCTCAGATCCCAAGCTCTCTCATTTATG 1356
|||||
451 GluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAlaThrLe 467
|||||
1357 GAAGCGGAACCTCTTAAAACTCTCTGATTTGAAGTTAGTACGCT 1406
|||||
467 uSerIleProLeuHisSerLeuAspThrGlyLysSerValThrIleHisA 484
|||||
1407 AAGTATTCCTCTCATTCCTTAGATAGTGAATAAAACCTCACTATCCACG 1456
|||||
484 laProAsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspGlu 500
|||||
1457 CCCCTAATCTTTCTATCCAAAAGATCTTCTCTTAAGCTGAGATGAG 1506
|||||
501 AsnPheTyGluAsnValGluLeuLeuSerLysGlyGlnAsnAsnIlePr 517
|||||
1507 AATTTTATGAATAATGTAGAGCTCTCAGTAAGACAAACAAATATTC 1556
|||||
517 oLeuLeuThrLeuSerLysGlyGlnSerHisLeuHisLeuProAspGlyA 534
|||||
1557 TCTCTTACTCTCCCTTAAAGACAAATCTCATTTACATCTCTCTGATGGGA 1606
|||||
534 snLeuSerSerHisPheGlyTyGlnGlyAspTrpThrPheSerTrpLys 550
|||||
1607 ACCTCTCTTCTACCTTGGATATCAAGGAGATTGACCTTTTCTTGAAA 1656
|||||
551 AspSerAspGluGlyHisSerLeuIleAlaAsnTrpThrProLysAsnTy 567
|||||
1657 GATTCGTGAAGAGGCGATCTCTGATGCTTAATTTGAGCGCCCAAAACTA 1706
|||||
567 rValProHisProGluArgGlnSerThrLeuValAlaAsnThrLeuTPA 584
|||||
1707 TGTGCTCATTCAGAACGTCATTAACACTCTTTCGAACACTCTTTTGA 1756
|||||
584 snTyLysSerAspMetGlnAlaValGlnSerMetIleAsnThrIleAla 600
|||||
1757 ACACCATTCGATATGCAAGCTGTGACACTGATGATTAATATCAACAGCG 1806
|||||
```


PT pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset
 PT asthma.
 XX
 PS Claim 16, Fig 1, 174pp, English.
 XX

CC This sequence is a Chlamydia antigen of the invention, designated
 CC CPN100634. The nucleic acids (and their complementary sequences) may be
 CC used as diagnostic agents for detecting the presence of nucleic acids
 CC encoding Chlamydia antigens in samples according to standard methods,
 CC and therefore, for diagnosing Chlamydia infections. For example, they may
 CC be used as primers and probes for diagnostic polymerase chain reaction
 CC (PCR) assays. Antisense sequences may be used to down regulate
 CC expression of the proteins and may be used to treat infections. The
 CC nucleic acids may also be used to produce the protein antigens they
 CC encode according to standard recombinant DNA methodologies. The
 CC proteins may then be used as antigens for the production of antibodies
 CC (i.e. as vaccines) for preventing infection by Chlamydia. The
 CC antibodies may also be used as diagnostic reagents for detecting
 CC infections. Chlamydia is a pathogen implicated in the development of
 CC (for example) community acquired pneumonia, upper respiratory tract
 CC disease (especially bronchitis and sinusitis, asthmatic bronchitis,
 CC adult-onset asthma and acute exacerbations of asthma in adults.

XX
 XX Sequence 928 AA:

Query Match 18.6%; Score 845; DB 21; Length 928;
 Best Local Similarity 28.8%; Pred. No. 3, 3e-62;
 Matches 247; Conservative 148; Mismatches 345; Indels 118; Gaps 26;

QY 97 DLSIQNRFSLFTDCSSKSSPS-11HOKNQLS-----LRNNGSMSECRNHAEGSCGA 149
 DB 113 nlftsgfslstf-----dssptlvltgqlgsagvnlrlrlvvaagntsladgga 166
 QY 150 ISADAFSLQHNLYFTAFEESSKNGGATGAQTF--LSRWSPISFARNADLNGAIC 208
 DB 167 lqgaftlltgsdgalftsmssstkggalaattagatamtyvtrflsmstssggald 226
 QY 209 SNLISGNVPLEFTGNSA-TNGACISIDLTSEKGS--LSLACNOETLFASNSAKEX 265
 DB 227 egsllsmnkfllyfegnaaktggalc-----nlkasgspellismnkllifsnvaeels 281
 QY 266 GCATYKAKHLYRNGPVSFINNSAKIF---GCAIAIQSGCSISLAGECVLFPONNSORT 321
 DB 282 ggaahkakkialssggtfelfrnnvssatpbgaisdaagelslaetgltfvrrlttl 341
 QY 322 --SDGLVNRATVLEKDALISLE-ARNDILFEDPIVQESSKESPESSLOASTSP 378
 DB 342 tgsctdprkmaingngkfetlraaknhtiffydpitsegssdvikingsagaalmpy 401
 QY 379 PATASPLVIOTSANRSVIFSSERLSEEE-KTPDNLTSOLOQPIELKSGRLVLDRAVLSA 437
 DB 402 qgt-----lltsgeltadelkvadnlksstfqpvsllgkylles 448
 QY 438 PSLISOPALLMEACTSIKTSS-DIKLATLISPLHSLDTEKSVTHAPNLSIQKLFISN 496
 DB 449 tsfsgaagslllgnsgsiltltagstlcnlgvdsllgkqvslakgas-nkvlvsg 507
 QY 497 -----SGDENFEVNLSEKEO-----NNPLPLTSLKSGSLILHLPDNLSS 537
 DB 508 klnlidiqniyes-hmfshdqlfslklitvdadvtnvdlssl-----lvpaeapds 560
 QY 538 HFGYQGDWTFMKNDSDEGHS--LIAMTPKNVPHPEROSTLVANTLMWYSDMQAVQSN 595
 DB 561 eyfgqgqwnvwtldatnltkeatwktqfvpersatvncnlwylfdirfqql 620
 QY 596 INTIAMGAVLGTWGSANVSNLFYAHDSGKPLIDNMHHHSLGLYLCFQISTHSLDHSFCLIA 655
 DB 621 veiqatqgmehkqgfvvssmnn--llhktgdenkkgfrhtsgvylvgssahltdddlflta 678
 QY 656 AGQLLCKSSDSFTSTETSYIATV-----QAQLATSLMK-----690
 DB 679 fchlrfardkdcflahnnstfyggclflfkshlclqpnylrlgraktsesaiackfpreipl 738

QY 691 -ISAQACYNESIHLEKTKYRSFSEKSGPSMHSVAASGEVCASIF-IVSNGSLFSSFSIF 748
 DB 739 alavqvsfnschirmethylst--pesegswsneclagglqldlftvlsnphlklftlpg 797
 QY 749 SKUGFSCTODGFESESGEIRSFSSFRNLSLPDITFEKKSOKRRTYVFLGAYIOL 808
 DB 798 mkveumyvsgnsfsssdgrfsigrllnlsipvgaekfvgglsydsydsfsvsdv 857
 QY 809 KRDVESGPVLLKNVSWDPAFMANLDSRAVMEFLTORALHRLQTLNNSC-----V 860
 DB 858 yrnmpstatlvmgsdsklrgnlsrgatllrgsn-----nyvynscellfghyame 910
 QY 861 LRQSHSYSLDCTFYRF 878
 DB 911 lrgstrnynvdygklrlf 928

RESULT 21

ID W88422 standard; Protein: 918 AA.
 XX
 AC W88422:
 XX
 DT 26-APR-1999 (first entry)
 XX
 DE Chlamydia pneumoniae surface exposed protein Omp9.
 KW Omp9, outer membrane protein 9; surface exposed protein; antigen;
 KW infection; diagnosis; vaccine; atherosclerosis; asthma.
 XX
 CH Chlamydia pneumoniae.
 OS W09858953-A2.
 XX
 PD 30-DEC-1998.
 XX
 PF 19-JUN-1998; 98WO-DK00266.
 XX
 PR 23-JUN-1997; 97DK-0000744.
 XX
 PA (BIRK/) BIRKELUND S.
 PA (CHRI/) CHRISTIANSEN G.
 PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
 PI Mygind P;
 XX
 DR WP1: 1999-105610/09.
 DR N-PSDB: X06821.
 XX
 PT Species-specific test for identifying mammals infected with
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
 PT these proteins
 XX
 PS Claim 7: Page 56-58; 115pp; English.
 XX
 CC This polypeptide comprises the novel 96.7 kDa surface exposed
 CC protein omp9 of the human respiratory pathogen Chlamydia
 CC pneumoniae. Its amino acid sequence was deduced from DNA (see
 CC X06821) isolated from a C. pneumoniae expression library. The
 CC invention provides 12 novel surface exposed proteins, Omp4-Omp15
 CC (see W88417-28), and nucleic acid sequences encoding them (see
 CC X06816-27). A new species specific test is claimed that is used
 CC to identify mammals (including humans) infected with Chlamydia
 CC pneumoniae. The test comprises detecting antibodies specific for
 CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
 CC membrane proteins, especially by PCR. The proteins are also used
 CC in the diagnosis of C. pneumoniae infection in mammals. The
 CC nucleic acids and proteins can also be used in the immunization of
 CC mammals, the nucleic acids being particularly useful as DNA
 CC vaccines for effecting in vivo expression of antigens. The
 CC vaccines may also prevent atherosclerosis and bronchial asthma,

OY 226 SAT-----NGAICISDLNTEKGSLSLACNOETLPFASNSAKEGCAIYAKKHVLRNCP 281
 Db 252 svlaetaggal-----sgdadvtlsqngsvtfsqngavangalYakKklllaag99 303
 OY 282 VS-----SAKIGAIATQSGSLSLAGEGVLFOONNS-QRTSOGGLVRNAIYL 333
 Db 304 ggnpfsmivgttaaggatlaagcslfseagqvlngnaivaltqgtkrmsidi 363
 OY 334 EKDALISLEARNG-DLFPDPVIOESSKESPLPSLSQASVTSPPATASPLVIQTSAN 392
 Db 364 gsvgtkdelraigsaffidpilaantaadstldlnkadedgn-----stdys 412
 OY 393 RSVIFSSERLSEER-KTPDNLTSQLAQPIELKSGRLVLRRAVLSASLSODQALLIME 451
 Db 413 gsvifsgklsedeakvadnltsllkqpvltagnlvkrgvltldtkgftqagssvilm 472
 OY 452 AGTSLKTS--DLKATLSIPLSLDTREKSVTIHAP-----NLSIOKIFLNSGDEMYE 504
 Db 473 agtlkasteavltlgisipvdsigegkvviaasakvnaalsgplllldng--naye 530
 OY 505 NVELLSKEON-----NIPULTISKEQSHLHPDGNLSHIFGOGDMFTSMKD---SDEGH 556
 Db 531 nht-lyktqdfstkvqalgaalcatctdvpaavpvauplhvgyqglvmtlwdtastpck 589
 OY 557 SLIANTPKNYVPHPERQSHLVANTLWNTYSDMQAVQSMINTIAHGAVALFTGWSAVSN 616
 Db 590 lallawenlgylnpergplvpnsllwgsstdqalqivlersalltcsdrgfwaagvan 649
 OY 617 LFAHDSGCRPIDMMHRSGLYLFISTHSDHSFCLAGQOLLGKSSDFITSTETSY 676
 Db 650 -fidkdkkgek-kyrkhsagyaigaaqtcsehlisafatqfsgskdftvaknhbdy 707
 OY 677 IATVOALATSL-----MKISAQCYNESIHEKTYRSPFSEKFG 717
 Db 708 agatylhilecsgfifcclldkipsqswshpvlvlegqlajshvsnldklylry-pek 766
 OY 718 SWSVAVSGEVCASIPVNSGSLFSSFSIFSKLQGFSGQDCESSGCEIRSFASSPFR 777
 Db 767 swgnafimmlgaashsypelyhctdyapyiklnlylqdsifexqtegrstdsni 826
 OY 778 NISLPITFEKSKQKRTTYFLGAIYQDLKRDVEGSPVLLKNVMDPMNLDSRA 837
 Db 827 nlslpivktekfscdndisyltisyvpllnpdkctalvisgawetayannlarga 886
 OY 838 YMFRLTQORALHRL-QTLNWSVLRGOSHSYSLDLGTTYRF 878
 Db 887 lqvtaagshyalspmfevlqgfvfevrgsrclynvdiqgkqf 928

RESULT 16
 W88429
 ID W88429 standard. Protein: 914 AA.
 XX
 AC W88429;
 XX
 DT 26-APR-1999 (first entry)
 XX
 DE Chlamydia pneumoniae surface exposed protein Omp5.
 XX
 KM Omp5: outer membrane protein 5; surface exposed protein; antigen;
 XX infection; diagnosis; vaccine; atherosclerosis; asthma.
 XX
 OS Chlamydia pneumoniae.
 XX
 PN MO9858953-AZ.
 XX
 PD 30-DEC-1998.
 XX
 PF 19-JUN-1998; 98MO-DK00266.
 XX
 PR 23-JUN-1997; 97DK-0000744.
 XX
 PA (BIRK/) BIRKELUND S.

PA (CHRI/) CHRISTIANSEN G.
 XX Birkelund S, Christiansen G, Knudsen K, Madsen A;
 PI Myglund P;
 XX WPT; 1999-105610/09.
 DR N-PSDB; X06828.
 XX
 PT Species-specific test for identifying mammals infected with
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
 PT these proteins
 XX
 XX Disclosure; Page 78-80; 115pp; English.
 PS
 CC This polypeptide is described as a subsequence of the claimed
 CC novel surface exposed protein Omp5 (see W88418) of Chlamydia
 CC pneumoniae, a human respiratory pathogen. The invention provides
 CC a new species specific test for identifying mammals (including
 CC humans) infected with C. pneumoniae. The test comprises detecting
 CC antibodies specific for surface exposed proteins Omp4-Omp5 (see
 CC W88417-28) or detecting nucleic acid fragments encoding them (see
 CC X06816-27), especially by PCR. The proteins are also used in the
 CC diagnosis of C. pneumoniae infection in mammals. The nucleic acids
 CC and proteins can also be used in the immunization of mammals, the
 CC nucleic acids being particularly useful as DNA vaccines for
 CC effecting in vivo expression of antigens. The vaccines may also
 CC prevent atherosclerosis and bronchial asthma, which are possibly
 CC associated with C. pneumoniae.
 XX
 SO Sequence 914 AA;
 Query Match 18 7%; Score 846.5; DB 20; Length 914;
 Best Local Similarity 30.6%; Pred. No. 2,4e-92;
 Matches 277; Conservative 137; Mismatches 383; Indels 107; Gaps 30;
 OY 12 CAALSTAVLFGQDPL-GETALLTKPNHVVVC---TFEEDCTMESTL--FPALCAHAS 63
 Db 20 cslvtaateenlgsdsfdsgntlgtlytkntltgldtltgltlgnlqdsaalckgcf 79
 OY 64 QD--DPLVYGNSTQWVSKLHTDPKEALPKKGLDISONFFPLSTDCSSKE--SSPS 119
 Db 80 sdtleslftagkygs--lsltnlkssaeq-----aalsvtcdlnsltgsfslftlaaps 132
 OY 120 II-----HKNQGLSLRNNGSMSECRNHAECSGASISADAFSLQHYLTAFEPENS 170
 Db 133 svltcpsgkavkcgqdlctdmngcllftkqdyceenggaatstknlsktsqsftegnk 192
 OY 171 SKG---NGAIOAQ-TPSLRNWSPISFANRRADLNGAI-CCSNLIGSGNVPLEFTGN 225
 Db 193 ssatgkkggalcalgtvdltnlntpltfsmnlaeaagallnsqnclltgnls-lyfsen 251
 OY 226 SAT-----NGAICISDLNTEKGSLSLACNOETLPFASNSAKEGCAIYAKKHVLRNCP 279
 Db 252 svlaetaggal-----sgdadvtlsqngsvtfsqngavangalYakKklllaag99 303
 OY 280 GPVSFFINN-----SAKIGAIATQSGSLSLAGEGVLFOONNS-QRTSOGGLVRNAIYL 333
 Db 304 gysifsmivgttaaggatlaagcslfseagqvlngnaivaltqgtkrmsidi 363
 OY 334 EKDALISLEARNG-DLFPDPVIOESSKESPLPSLSQASVTSPPATASPLVIQTSAN 392
 Db 364 gsvtkltnlraigsaffidpilaantaadstldlnkadedgn-----stdys 412
 OY 393 RSVIFSSERLSEER-KTPDNLTSQLAQPIELKSGRLVLRRAVLSASLSODQALLIME 451
 Db 413 gsvifsgklsedeakvadnltsllkqpvltagnlvkrgvltldtkgftqagssvilm 472
 OY 452 AGTSLKTS--DLKATLSIPLSLDTREKSVTIHAP-----NLSIOKIFLNSGDEMYE 504
 Db 473 agtlkasteavltlgisipvdsigegkvviaasakvnaalsgplllldng--naye 530

in the diagnosis of *C. pneumoniae* infection in mammals. The nucleic acids and proteins can also be used in the immunization of mammals, the nucleic acids being particularly useful as DNA vaccines for effecting in vivo expression of antigens. The CC vaccines may also prevent atherosclerosis and bronchial asthma, which are possibly associated with *C. pneumoniae*.

Sequence 928 AA:

Query Match 18.8%; Score 857; DB 20; Length 928;
Best Local Similarity 30.2%; Pred. No. 3, 2e-63;
Matches 261; Conservative 123; Mismatches 367; Indels 114; Gaps 24;

```

74 SYCMFVSKLHITDPKELFEKGLDSTQNFRLSFTQSSKSSPSIIHQKNGQLSRNN 133
Db 118 tllqfssistlaspgssltqkgav-----cst-----gslkfdkn 154
QY 134 GSNMFCNNHAEAGCAISADAFSLQHNLYLFAFEENSSKNGCAIQ-AQFSLSRNVSPI 192
Db 155 vslltskftdngaltakltlslgtlmsallsentsskkgalqtsdaltlqngqv 214
QY 193 SFARNRADLNGALCCGNCNLGSGNVPLEFTGSAT-----NGAICISDLNTS 242
Db 215 sfadtsdsqaaiftleasvltlnakvstfidhkvltgassstlqdmvgalca---ykt 271
QY 243 EKGSLSLACNGEPLFASNAKEKGATYAKHMYLRNGPVSFINNS-----AKIGAIAT 297
Db 272 tdkvtlltgmllfsmntctaggaitykkliaasglllfsnsvngctabkggatal 331
QY 298 QSGGSLSLIAGEGSLVLPNNNSORTSDGLVMAVLEKDAISSLKLRNG-DILFEPD 356
Db 332 edegelsladsdglvltgtvtstlprlrrsldltskmlalraagraalfyfpilc 391
QY 357 QESS-----KESPLPSLSQASVSPTRATAPLVIOTSAKRSVFSSERISEEKT 408
Db 392 tgsstvtvdkvneprdsalqyt-----gnllfeyeklselea 432
QY 409 PD-NLTSQLQPIELKSGRLVLDRAVLAPSLSDQPALIMAGSLTSSDLKATL 467
Db 433 dsknltskllqprlsgltslkhyvltlqatfqgadsrlendvgltlepadstlnl 492
QY 468 SLPRLSLDTEKSVYIH---APNLISLQKIFLSNSGDEFFYENWELLSKONNIP 522
Db 493 vlnissidgakkietkatskntlsgltllldplgtfynhlnrpsydlelkasg 552
QY 523 KEQSHLPLPGNLSS--HFGYQGM--TFMKDSDEGSLA--NMTFRNVPVHERQSTL 577
Db 553 tvstlvtppdlmgekfhygyqgtwpiw---gtgastlatlwtktgyipmperigs 609
QY 578 VANTLNTYSDMOAVQSMINTIAHG--GAYLFGTNGSVAVNLFEYAHDSGKPTDNM 635
Db 610 vpslwnafidissihymetaneqlqdrat--wcaglsnlt--hkdsklrrgtfnhs 665
QY 636 LGYLFGLSTHLDHDFCLACQLLCKSSDPSFTSTFTSYIATVQAQLATSLMKI--- 691
Db 666 ggyvltggnlhtcskllsaafcgfgrdtyfvakngqvtvyggllyqhmetylsdck 725
QY 692 -----SAQACYNESIHLEKTKYRSFSEKFGSHVAVGEVCASTPIV 735
Db 726 rpslslyvpeipvlflsgnlsyhtdndlktkyltptvk-gsvngndsfalefgripic 784
QY 736 SNGGGLFSSSIFSKLOGFGTGDFEESGSEIRSFSSFRNLSLPIGTFFEKSG-KT 794
Db 785 ldeaalteqymplmkifvyahneglkeqglearefsgslvnlalrldfdesdcd 844
QY 795 RYYVYFLGAVYIDLKRDVEGSPVLLKNNAVSWAPMANLDSKRAYMFLRQNALH-RLQ 853
Db 845 atynltlg-ytvdvrsnptcltllrsgdsklfgnlarqalvllagphfcfnflea 903
QY 854 LNAVSVLRGSHSYSLDLGTTTFE 878
Db 904 fsqfsltgstsrnyndvlqakgkf 928

```

RESULT 14

ID Y69368 standard: Protein; 945 AA.

AC Y69368;

DT 19-JUN-2000 (first entry)

DE Amino acid sequence of the CPN100394 polypeptide.

KW CPN100394; Chlamydia infection; Immune response; vaccine.

OS Chlamydia pneumoniae.

FH Key Location/Qualifiers

FT Peptide 1..43 /note="signal peptide"

PN W020001183-A2.

PD 02-MAR-2000.

PF 18-AUG-1999; 99MO-IB01449.

PR 20-AUG-1998; 98US-0097187.

PR 20-AUG-1998; 98US-0097188.

PR 20-AUG-1998; 98US-0097189.

PR 20-AUG-1998; 98US-0097190.

PR 20-AUG-1998; 98US-0097195.

PR 20-AUG-1998; 98US-0097196.

PR 20-AUG-1998; 98US-0097197.

PR 27-AUG-1998; 98US-0097191.

PR 17-AUG-1999; 99US-0376770.

PA (CONN-) CONNAUGHT LAB LTD.

PI Murdin AD, Oomen RP;

DR WPI; 2000-224703/19.

DR N-PSDB; Z61508.

XX Novel antigens and corresponding DNA molecules that can be used to

XX prevent, treat and diagnose disease caused by Chlamydia infection in

XX mammals, especially humans -

XX Claim 18; Fig 13A-F; 20pp; English.

PS Y69362-69 represent Chlamydia pneumoniae polypeptides. The polypeptides

XX are present in the bacterial membrane structure, in the external

XX vicinity of the membrane structure, in the inclusion membrane

XX structure, in the external vicinity of the infected cell. The polypeptides may be

XX used to prevent, treat and detect the presence of Chlamydia infection

XX and/or the presence of Chlamydia in a sample. The polypeptides may

XX also be used to induce an immune response in a mammal. The vaccine

XX vector comprising the polynucleotides is used to induce an immune

XX response in a mammal. Antibodies directed against the polypeptides

XX may also be used therapeutically to treat and/or prevent a Chlamydia

Sequence 945 AA:

Query Match 18.8%; Score 851; DB 21; Length 945;
Best Local Similarity 28.6%; Pred. No. 1, 1e-62;
Matches 248; Conservative 153; Mismatches 367; Indels 98; Gaps 25;

```

74 SYCMFVSKLHITDPKELFEKGLD-----STQNFRLSFTQSSKSSPSIIHQKNGQ 127
Db 117 sfdlslvtspksavttcgslvsglvgvldntlvltsnasvedgvi--kynsc 174

```

ID Y94327 standard; Protein: 928 AA.
XX
AC Y94327;
XX
DT 11-AUG-2000 (first entry)
XX
DE Chlamydia pneumoniae 98kD putative outer membrane protein.
XX
KW Chlamydia; antigen; vaccine; infection; outer membrane protein.
XX
OS Chlamydia pneumoniae.
XX
PN MO200026237-A2.
XX
PD 11-MAY-2000.
XX
PE 29-OCT-1999; 99WC-GB03579.
XX
PF 29-OCT-1998; 98US-0106070.
XX
PR 01-MAR-1999; 99US-0122066.
XX
PR 27-OCT-1999; 99US-0428122.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Mordin AD, Oomen RP, Dunn PL;
XX
DR MPI: 2000-365569/31.
XX
DR N-PSDB: A27021.
XX
PT Novel Chlamydia 98 kDa putative outer membrane protein antigen, used
XX for vaccination and protection against Chlamydia infection .
XX
PS Claim 6; Fig 1; 93pp; English.
XX
CC The present sequence is the 98kDa putative outer membrane protein from
CC Chlamydia pneumoniae. The genomic sequence was amplified using two
CC PCR primers. The 5' primer contains a NotI restriction site, a ribosome
CC binding site, an initiation codon and a sequence close to the 5' end of
CC the 98kDa putative outer membrane protein coding sequence. The 3' primer
CC contains the sequence encoding the C-terminal sequence of the putative
CC outer membrane protein and a BspGI restriction site. The stop codon was
CC excluded and an additional nucleotide was inserted to obtain an in-frame
CC C-terminal fusion with the histidine tag. The PCR product was cloned
CC into a eukaryotic expression vector (pCA-Myc-His) by restricting both
CC the vector and the PCR product with NotI and BamHI and performing a
CC ligation reaction. This expression vector was injected intramuscularly
CC and intranasally into mice, which were subsequently inoculated with
CC Chlamydia pneumoniae. The chlamydial lung titers of the immunised mice
CC were lower than those of the controls. Thus the 98kDa putative outer
CC membrane protein can be used as a vaccine to provide protection against
CC Chlamydia infections, especially Chlamydia pneumoniae infections.
CC The present polypeptide may also be administered orally to treat
CC Chlamydia infection.
XX
SQ Sequence 928 AA:

Query Match 19.1%; Score 866; DB 21; Length 928;
Best Local Similarity 30.4%; Pred. No. 5 6e-64;
Matches 263; Conservative 133; Mismatches 365; Indels 114; Gaps 24;

QY 74 SYCMFVSKLHITDPKALFKKGLDLSIONFRPLSTFDCKSSKSSPSIIHOKNQLSRNN 133
DB 118 LFIgfssLFIaapsalltgkavs-----cst-----gslsltkn 154
QY 134 GMSFPCNHAEGSGAISAFAFLQHNHYLFTAFENSSKNGALQ-AQFFSLSRNVSP 192
DB 155 vsllfskntscdnggatactislctsmaltsemsskkgalqtsdaltltgnygev 214
QY 193 SFARNRADLNGAIGCSNLTICSGVNPDLFTGNSAT-----NGAICICISDLNLS 242
DB 215 sfadntssdsagaiftcaasvltisnmakvstfdhkvvgasssttgdmssgysaice---ykt 271

QY 243 EKGSLSLACNQETFLFASNSAKEKGAIYAKHMVLRVNGVSPFINNS-----AKIGCALAI 297
DB 272 tdtkvltlqgmllfennsttagaiyvkkllelaeggltilfsmnsvngtarpkgalai 331
QY 298 QSGGSLIILAGESSVLFQNNNSQRTSDOGLVRNAYILEKNAIISLSLEARNG-DILFFDP 356
DB 332 edsgeislssadsqdvlfngntvsttptgntsrldgtsakmtalrsaagralyfydpt 391
QY 357 QESSS-----KESPLPSSLOASVTSPTPATASPLVITQTSANRSVFSRSLSEEXT 408
DB 392 tgsstvtvklvnefpadaalqyt-----gnlltfgklsereaa 432
QY 409 PD-NLTSQIQPIELKSGRLVLRKRAVLAPSLSODPOLIMEACTSLKTSDDLKATL 467
DB 433 dsknltskllqpvlsqgltslkhyvltlqgafitqgaderslmdvytllpepdtstlnl 492
QY 468 SIPHLSDTEKSVTH-----APNLSTQKIFLSNGSENYENVELSKQNNIPLTLS 522
DB 493 vniissidgakkakietaktsknltlsgtcltlpctgyenhslrnpqsydllelkaag 552
QY 523 KEQSHLHPDGNLSS--HFQYQGDW-TFSMKDSDEGSLIA--NMTPKNYVPHPERQSTL 577
DB 553 tvstslavtpdpimgekfhygygtwpglvw---gtgaatcatcnvcktgylmpetigs 609
QY 578 VANTLWNTYSDMQAVQSMINTIAHG--GAYLFQTMGSAVSNLFYAHDSGKPIDMNHRS 635
DB 610 vpsnlwnafldisslhyImetaneigqgdraf--waglsnff--hkdstkrrgrfhis 665
QY 636 LGTLFGLSTHSLDDHSPCLAAQGLLCKSSDFITSTETSYIATVOAQLATSLMKI---- 691
DB 666 ggyvlygnhlhtcsdkllsaefcqlfgdrdyfvakqgtvlyggtlyghneyislpc 725
QY 692 -----SAQACYNESIHELKTKYRSRSGKGFSGMHSVAVSGEVCASIP 735
DB 726 rpsclsyvptelpvltsnglsyhtndlktlytlypvtk-gswgndstalefgtgrapic 784
QY 736 SNGSGLFSSFSIFSKIQGFSGTODGFESSGGEIRFSASFRNISLUPIGITEKKSQ-KT 794
DB 785 ldsaalfeqymplmkqfyvahqegfkeqgtearefsgsrlnvialpigrfdkeedq 844
QY 795 RTYYRFLGAYIDOLKRDVESGPVYLKNAVSWAPRANLDSRAYMRRLNQRALH-RLOT 853
DB 845 atynlclg-yvdlvtrspndcttllfsgdswktlgtlnlargalvtragnhtcfnsfnea 903
QY 854 LNWSCVLRQSHSYSLDGLTTRYF 878
DB 904 fsgfslrfgsrnynvdigakqgf 928
RFSULT 12
Y69369
ID Y69369 standard; Protein: 918 AA.
XX
AC Y69369;
XX
DT 19-JUN-2000 (first entry)
XX
DE Amino acid sequence of the CPN100395 polypeptide.
XX
KW CPN100395; Chlamydia infection; immune response; vaccine.
XX
OS Chlamydia pneumoniae.
XX
PN MO200011183-A2.
XX
PD 02-MAR-2000.
XX
PF 18-AUG-1999; 99WC-IB01449.
XX
PR 20-AUG-1998; 98US-0097187.
XX
PR 20-AUG-1998; 98US-0097188.
XX
PR 20-AUG-1998; 98US-0097189.
XX
PR 20-AUG-1998; 98US-0097190.


```

Db      349 nltststapstrnalytgsakltmraagqsiyfydplaantgasvlt----- 400
Qy      375 TSPPTATASPLVYIQTSAKNSRVIFSSERLS--EEKTPDNLTSQLOQPIELKSGRLVLDRA 433
Db      401 tlnqpdnsnpl-----dysgltvifsgklsadeakaadnftslkqplalasgtlalkgnv 456
Qy      434 VLSAPSLSDPPQALLIMEAGTSUKTSSD-LKLATLSLPLSLDMEKSVTHAP-----N 486
Db      457 eldvngftqlegstllmqpgtklkadeaislklvldlsalegnksvstetaganakltc 516
Qy      487 LSIQKIFLNSGDENFYENVELLSKEQNNIPLLTLSKEQSHLHPDGNLSS-----HF 539
Db      517 ltlpvlvfydssg--nfyshtlmgafcpqlvvtfaataasdiyl-dalltspqtphepy 573
Qy      540 GYCGDMTFESKSDSEGLSLANMTPKNYVPHPEROSTLVANTLWNTYSDMQAVOSMINTI 599
Db      574 gygghweatwadtsaksgtmctwtgynperrasvdpdsiwasftdlrtlqintsg 633
Qy      600 AHGCAVLFCTWGSVSNLFLFAHDSGKPIIDNMHHRSLGYLFGISTHLDHSFCLAGOL 659
Db      634 anslyqgrqlwaagtanfth-kdksqtn-qaftrkksygylvgsaaedfsenlfsvalcql 691
Qy      660 LCKSSDSFTSTETTSYIATVQAO-----LATSLMK-----ISAQACYNES 700
Db      692 fykdldlftvensshyjaelyqhrafqglmpstfgstlmdlkdplilnaglsysyt 751
Qy      701 IHEKTRYSRFSKEGFSMSHVAVSGEVCASIPV-VSNGSGLSSFSFESKLGQFGSGTOD 759
Db      752 kmnditrysy-peaagswtensgalelgyssaiylpkcapffqgyfpfklfqavysrty 810
Qy      760 GFPESSSEISFSSASSFRNLSLPIGTFEKKSKOKTRFYFFYFIAGYIODLKRDVESCHVVL 819
Db      811 nfksgaeaaftddgdlnvcsipqitlekiseekmfelstlayldqvyrknprrtsl 870
Qy      820 LKNAVMDAPMANLDSRAVFRLTNORALH-RLQTLNVSCVLKQSHSYSLDICTTTRF 878
Db      871 mvsgaststclcnlargaftasagshltlsphealsgeaayeltgsahlynvdcgllysf 930

RESULT
ID      9
W88424 standard; protein; 930 AA.
XX
AC      W88424;
XX
DT      26-APR-1999 (first entry).
XX
DE      Chlamydia pneumoniae surface exposed protein Omp11.
XX
KM      Omp11: outer membrane protein 11: surface exposed protein;
XX      antigen; infection; diagnosis; vaccine; atherosclerosis; asthma.
XX
OS      Chlamydia pneumoniae.
XX
PN      WO9858953-A2.
XX
PD      30-DEC-1998.
XX
PF      19-JUN-1998: 98WO-DK00266.
XX
PR      23-JUN-1997: 97DK-0000744.
XX
PA      (BIRK/) BIRKELUND S.
XX      (CHR/) CHRISTIANSEN G.
XX
PI      Birkelund S, Christiansen G, Knudsen K, Madsen A;
XX      Mydgin P;
XX
DR      WPI: 1999-105610/09.
XX      N-PSDB: X06823.
XX
PT      Species-specific test for identifying mammals infected with
```

Chlamydia pneumoniae - comprises detecting antibodies specific for outer membrane proteins of C. pneumoniae or nucleic acids encoding these proteins

Claim 7: Page 63-65; 11pp; English.

This polypeptide comprises the novel 97.6 kDa surface exposed protein Omp11 of the human respiratory pathogen Chlamydia pneumoniae. Its amino acid sequence was deduced from DNA (see X06823) isolated from a C. pneumoniae expression library. The invention provides 12 novel surface exposed proteins, Omp4-Omp13 (see W88417-28), and nucleic acid sequences encoding them (see X06816-27). A new species specific test is claimed that is used to identify mammals (including humans) infected with Chlamydia pneumoniae. The test comprises detecting antibodies specific for Omp4-Omp13 or detecting nucleic acid fragments encoding these outer membrane proteins, especially by PCR. The proteins are also used in the diagnosis of C. pneumoniae infection in mammals. The CC nucleic acids and proteins can also be used in the immunization of mammals, the nucleic acids being particularly useful as DNA vaccines for effecting in vivo expression of antigens. The CC vaccines may also prevent atherosclerosis and bronchial asthma, which are possibly associated with C. pneumoniae.

Sequence 930 AA;

Query Match 20.0%; Score 906.5; DB 20; Length 930;
Best Local Similarity 30.7%; Pred. No. 2,1e-67;
Matches 258; Conservative 150; Mismatches 345; Indels 87; Gaps 26;

```

Qy      98 LSIQNFHPLFSF-----TDCSSKSSPPIIHQKNGQLSLRNGSMSPCR--NHAEGSCGA 149
Db      119 lltfglnsfliaapgtlvtagksltls-----sagalltdnglllfsqvnseannmgya 174
Qy      150 ISADAPSLQHNLYLFTAEENSCKNGCAI-OAQFTLSRNVSPISFARNRADLNGAICG 208
Db      175 ltkcltislsgtststlftnsakllygaelyssaaslsqgtglvfmnnkgeggagll-- 232
Qy      209 SNLICSGNV---NPLFTGNSATN---GGAICISDLNTSEKGSLSLACQETLRASNS 261
Db      233 -gfeassstlcqnsllffsgntatdaagkgaylc--ekgetpdltlisgnkslftaens 288
Qy      262 AKEKCAIVAKHMYLRYKNCVVFINN-----SAKIGCALIOSGSGSLIAGGSYLFON 316
Db      289 svcggaicahgldisaagpdlfsmrcgntaagkgajaladagslsisangddlftlg 348
Qy      317 NS-ORTSDQGLVRNAYILEKDAISSLARNG-DILFPDPIVOESSSEKSPPLSSLOASY 374
Db      349 nltststapstrnalytgsakltmraagqsiyfydplaantgasvlt----- 400
Qy      375 TSPPTATASPLVYIQTSAKNSRVIFSSERLS--EEKTPDNLTSQLOQPIELKSGRLVLDRA 433
Db      401 tlnqpdnsnpl-----dysgltvifsgklsadeakaadnftslkqplalasgtlalkgnv 456
Qy      434 VLSAPSLSDPPQALLIMEAGTSUKTSSD-LKLATLSLPLSLDMEKSVTHAP-----N 486
Db      457 eldvngftqlegstllmqpgtklkadeaislklvldlsalegnksvstetaganakltc 516
Qy      487 LSIQKIFLNSGDENFYENVELLSKEQNNIPLLTLSKEQSHLHPDGNLSS-----HF 539
Db      517 ltlpvlvfydssg--nfyshtlmgafcpqlvvtfaataasdiyl-dalltspqtphepy 573
Qy      540 GYCGDMTFESKSDSEGLSLANMTPKNYVPHPEROSTLVANTLWNTYSDMQAVOSMINTI 599
Db      574 gygghweatwadtsaksgtmctwtgynperrasvdpdsiwasftdlrtlqintsg 633
Qy      600 AHGCAVLFCTWGSVSNLFLFAHDSGKPIIDNMHHRSLGYLFGISTHLDHSFCLAGOL 659
Db      634 anslyqgrqlwaagtanfth-kdksqtn-qaftrkksygylvgsaaedfsenlfsvalcql 691
Qy      660 LCKSSDSFTSTETTSYIATVQAO-----LATSLMK-----ISAQACYNES 700
```


OY 367 PSSLIQASVTSPPATASPLVIQTSANRSLVFSSEERLSEEEKTPDNTLSQLOQPIELKSGR 426
DB 339 PSSIGASVTSPTAPASPLVIQTSANRSLVFSSEERLSEEEKTPDNTLSQLOQPIELKSGR 398
OY 427 LVLKDRVLSAPSLSDQPPQALLMEACTSLKTSDDLK 464
DB 399 LVLKDRVLSAPSLSDQPPQALLMEACTSLKTSDDLK 436
RESULT 4
ID Y90239 standard: Protein: 928 AA.
AC Y90239;
XX
DT 29-AUG-2000 (first entry)
XX
DE Chlamydia antigen CPN100638.
XX
KM Chlamydia antigen; diagnosis; infection; community acquired pneumonia;
KM therapy: upper respiratory tract disease; bronchitis; sinusitis;
KM asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma.
OS Chlamydia pneumoniae.
XX
PN WO200032794-A2.
XX
PD 08-JUN-2000.
XX
PF 01-DEC-1999: 99WO-CA01147.
XX
PR 01-DEC-1998: 98US-0110339.
PR 01-DEC-1998: 98US-0110340.
PR 01-DEC-1998: 98US-0110427.
PR 01-DEC-1998: 98US-0110428.
PR 01-DEC-1998: 98US-0110438.
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Murdin AD, Oomen RP, Wang J;
XX
DR WPI: 2000-412339/35.
XX
N-PSDB: A30851, A30852.
PT
PT Nucleic acids encoding polypeptide antigens from Chlamydia useful for
PT preventing, diagnosing and treating diseases such as community acquired
PT pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset
PT asthma.
XX
PS Claim 16; Fig 5; 174pp: English.
XX
XX This sequence is a Chlamydia antigen of the invention, designated
CC CPN100638. The nucleic acids (and their complementary sequences) may be
CC used as diagnostic agents for detecting the presence of nucleic acids
CC encoding Chlamydia antigens in samples according to standard methods,
CC and therefore, for diagnosing Chlamydia infections. For example, they may
CC be used as primers and probes for diagnostic polymerase chain reaction
CC (PCR) assays. Antisense sequences may be used to down regulate
CC expression of the proteins and may be used to treat infections. The
CC nucleic acids may also be used to produce the protein antigens they
CC encode according to standard recombinant DNA methodologies. The
CC proteins may then be used as antigens for the production of antibodies
CC (i.e. as vaccines) for preventing infection by Chlamydia. The
CC antibodies may also be used as diagnostic reagents for detecting
CC infections. Chlamydia is a pathogen implicated in the development of
CC (for example) community acquired pneumonia, upper respiratory tract
CC disease (especially bronchitis and sinusitis), asthmatic bronchitis,
CC adult-onset asthma and acute exacerbations of asthma in adults.
XX
SQ Sequence 928 AA:

Query Match 20.8% Score 944.5; DB 21: Length 928;

Best Local Similarity 31.6%; Pred. No. 1,3e-70;
Matches 288; Conservative 137; Mismatches 336; Indels 149; Gaps 32;
OY 85 TDPEALFKRKQDLSIONFRF-LSFTDCSSKESSPSLHKKNG-QLSLRN----- 132
DB 52 tnadgtynlvgdvsitnagspatascfctetgnlslfghyqglfllnqidagancftc 111
OY 133 -----NCMSRF-----CRNHAEGSGCAISADAFSLQHNLYFTAFENSSKGNCGAI 178
DB 112 ntaanklslsfsgfsysllcltnatcgtgalktgcacslsny-scylqgnfndnggal 170
OY 179 QAOFTSLSRNVSPISFARNRADINGAICCSNLICSGN-VNPLFTFGNSATN----- 229
DB 171 qgsstislhnp-ltfaknktqkgalytggtltnntlnasfsentannnggalyte 229
OY 230 -----GAIICISDLNINSEKSLSLACNOETLFXNSAKEX 265
DB 230 asffisnkaistinnsvtatsatgalyccs--tsapkvltlsdngelnfignatals 287
OY 266 GCAIYAKHWLRYNCPVSPFINNS-----AKIGAIATIOSGSLSLIAGEGSLVFOUNNS-- 318
DB 288 ggalycdnhlvylssgprllfnnsgydaaplgaialadsgslsialagdlfegntlv 347
OY 319 -QRTSDGGLVRNAIYL-EKDAIILSLEARNGD-ILFEDPIVOESSKESPLPSLQASVT 375
DB 348 kgasssgttrnslnigntnakivqlraagntilyfydpt-----tsitaals 396
OY 376 SPTPATASPLVIQTSANRSLVFSSEERLSEEEKT-PMULTRLSQLOQPIELKSGRLVKDRAY 434
DB 397 dalnlgpdlnagpnayggtlvfsgkiseaaadlktlqptltaggqlslksyvc 456
OY 435 LSAPSLSDQPPQALLMEACTSLKTSDDLKATLSIPRLSDTKESKYTHAPNLSLQIFL 494
DB 457 lvakstfsgpsstllmdagtlleadgdltnlnlvndskelkgtlktqgs-qtvl 515
OY 495 SNS-----GDENFEYENVLLSKDONN--IPLLTSLKEQ-SHLNLPDGNLS----- 537
DB 516 sgslslvdpsgnyvedy-----swnpqvfcsltleadpanhltd--laadpleknp 568
OY 538 HFGYQGDMTFSWKDSDGCHLLA--NMTPKNYVPHRROSTIVANLTMNYSDMOAVQSM 595
DB 569 hwygqgnwalswgedtaekkaatlwtcktgynpneritglvantlwgafvdrisq 628
OY 596 INTIAHGAYLFGTWSAVSNLFYAHDSCKPIDNMHNRLGVLFGISTHSDHSCCLA 655
DB 629 valkvrgsgelrglweaglnft--hkdstklkghfhisagylvvgatcllaadnlita 686
OY 656 AQOLLKSSDSFTSTETTYIATVQAQ-LAT-----SLMK-----ISAQACYN 698
DB 687 fcqffgkrdhflnknaasayaaashlqhlaltsspslllylpyseseqpvlfdaqsyl 746
OY 699 ESTHELKTXYRSPKBEFGSMHGVAVSGEVCASIPYV-NGSLGFSFSIFSKLQGFSGT 757
DB 747 yskntcmtytqapk-gesswynygcalelsslphlalsheglfmayfrfkveasyih 805
OY 758 QDSFEER-SSGEIISFSASFRNISLPIGIFFEKSKOKTRTYVFLAYIIDLKRVESGP 816
DB 806 qdsfkenltllvsldsgdlinvsrpilgiferfimeretsyaavlyvadvyrknpdct 865
OY 817 VLLKNAVSWDAPMANLDSRA-----YMFRLTNORALHQLTLLNNSCVLRGSHSY 868
DB 866 talllnntskctgtlnlsrqaglygraglfyafs-----pnlvetslnsmelgrsrsy 918
OY 869 SLDLGTYRRF 878
DB 919 nadlgykftf 928
RESULT 5
ID W88423 standard: Protein: 928 AA.
XX
AC W88423;

Query Match 2.8%; Score 129; DB 2; Length 1600;
 Best Local Similarity 18.3%; Pred. No. 0.021;
 Matches 149; Conservative 107; Mismatches 275; Indels 284; Gaps 36;

73 NSYCFVSLHTDPKELFKK-----DLSTIONFRFLSTDCSSKESPSIIHOKNG 126
 143 NOCVFLINMGITICKDALINTNGFTASTLDSINENIKARNTLQOTDKALAEI-VNHG 201
 127 QLSLNNNSMFCRHHAECSGCAISADAFSLQHNLYLFAFEENSSKNGCAIQ---AQTF 183
 202 LITVCKDQSVNLI-----GGKVK-----NEGVIASVNGSISLLAGQKI 239
 184 SLGRVSP-----ISFARRADLNGALICCSNLICSGVNNPLFPNGNSATNGALICISD 238
 240 TISDIINPTITYSIAAPENEA-INLQDIFAK-----CGNIN-----VRA 277
 239 LNTSEKSLSLACNOETLFASSAKKGAAY--AKHMYLRNGPYFINSNAK----- 290
 278 ATIRNKGKLS-----ADSVSKDKSGNIVLSAKEGAEIGCVISAONOAGKCKLMI 328
 291 ICGAIAIQSGSLSLAGEGCVI-PONNSQRTSDQGLVNAIYLEKDALLSLEARN----- 346
 329 TQDKVTLKGAVIDLSGREGGYLGGDER-----GEGKNGIQAKKTKLEKSTINVSCK 384
 347 -----GDLFPPIVQESS-----361
 385 EKGRATVWCIDIALIDGNINAQSDIAKTGCFVETSGHOLSTICDDVIYDAKEWLLDDPDV 444
 362 -----KESPLPSLQASVTSPPTASPL-----VIQT 389
 445 SLETTLSGRNNGENOCYTTCOTKRSPPKN-----SISKPT-LTNSTLQDLRGSYVNI 499
 390 SANRSVIFSSERLSEBEKTPDNLTSOLOQPIELKSGRLV--KDRVLSAPSLSDPOL 447
 500 TANNRIYVNS-----INLSNGSLTLHTRKQDVKINGDITSENNGN 540
 448 LINEACTSLKSSDLKATLSTPLHSIDT-----EKSTTHAPNLSIQKILPSNGDE 500
 541 LTKAKSMDVHNKNTLTCGFLNIVAGDSVAFEREDKANATDAOITTAQCTTVKNDK 600
 501 NF-YENVLLSKFONNPLTLTSEKSHLILPDGNSLHSHFGYOGDWTFS-----WK 550
 601 QFFNNVSLNGCTGK---LKFIANQNNFTHKPIGCEIN---ISGIYIINTTKKDYWN 653
 551 DSDGHSLSLANMTPKNVYPHEROSTLVANTLM-----NTYSDMQAVQSMJNTIAHG 602
 654 ASKDSY-----W-----NVSLLTNTYQKTFIKFVDSGSMQODLRSSRSFA-- 696
 603 GAVLFGTWSAVSNL-----FYAH-----DSSCKPIDNMH 632
 697 CVHFNGICGKTNNFICANAKALFKLNATDPKELPITFNNANITATVGCNSDSSVMEFDIH 756
 633 HRSGLVFGTSTHSDHSCFLAAGOLGKSSDSFSTSTETSYIATVQALATSLMKIS 692
 757 ANLTSRAAGINMDSIN-----ITGGL-----DESIISHMKNNAFIRKIDLTT---IN 800
 693 AOCYNESIHELKTR-YRSPSKGEGFSGMBSAVASG 726
 801 ATGS-NFSLKOTKDSFYNEYSKHAHNSHNLITLG 834

CITY: San Francisco
 STATE: California
 COUNTRY: United States
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/10661A
 FILING DATE: 16-AUG-1995
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/296,791
 FILING DATE: 25-AUG-1994
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Treccarlin, Richard F.
 REGISTRATION NUMBER: 31,801
 REFERENCE/DOCKET NUMBER: FP-59941/RFT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELEX: 910 277299
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1702 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
 PCT-US95-10661A-5

Query Match 2.8%; Score 127.5; DB 5; Length 1702;
 Best Local Similarity 20.1%; Pred. No. 0.033;
 Matches 182; Conservative 119; Mismatches 296; Indels 309; Gaps 50;

66 DPL---YVLGNSYCFVSKLHTDPKELFKKGD-LSTIONFRFLSTDCSS----- 113
 277 DPLTNVAVLGD---GSLPYVD-----REKGMVLPFGSYFMAVGMKMSQENNIYK 326
 114 -----KESPSIITHOKNGOLSLRNNGSMSFCRHHAECSGCAISADAFSLQHNLYLT 164
 327 PEFKATVLDKDTAGSLI-GSNTQYNNNPTGKTSVISNGSE---SLNVDFD-----S 374
 165 AFFENSSKNGCAIQNOTFELSNNVSPISFARRADLNG-GAICCSNLICSGVNNPLFFT 223
 375 SQDTDSKKNHG-----KSVTLRSGGTLTNNNIDOG-AGGLFEE 413
 224 GNSATNGAICISDINTSEKSLSLACNOETLFASSAKKGAAYAKHMYLRNGPYV 283
 414 GQYEVKG-----TSDSTTWKAGAVSA-----DQKTVTKVHNPKS 449
 284 FINNSAKTGAIAIQG-----GSLTLAGEGYLFO-----NNSORTSDQ-GLV--RNA 330
 450 --DRLAKIGKGLIVEGKNGKSLKV--GDGVYILKQADANNKYVAFSQVQIVSGRST 505
 331 IYLEKDAIIL-----SLEARNGDILFPDPI-----VOESSKESPLPSL 370
 506 VVLNDDKQVDPNNTSYFGFGRGRDLA--NGNNLTFEHIINIDGARYVHNNTSKSTVYITG 564
 371 QASVTSPTPATASPLVITQTSANRSVIFSSERLSEBEKTPDNLTSOLOQPIELSGRLVK 430
 565 ESLITDENTY-----TPYN-----578
 431 DRAVLAPSLSODPOLLINEAGTSLKTSDDLKATLSTPLHSIDTEKSYTHAPNLSIQ 490
 579 ----IDAPD-EDNRYAF-----RRIKDGGLYLNLENTYVYALRKGASTRSELP----- 622
 491 KIFLSNGDENFENVNVELLSKQONNPLTLTSLKQSHLPLDGVLSHFGYOGDWTFSWK 550
 623 ---KNGGESH--ENMLYMKTSD-----AAKKNVNMHNINEMNNGFNIGTGE----- 664

QY 462 LKATLSIPLHS---LDTEKSV---THAPNLSIOKIFLSN-----SCDENFENVELLSKE 512
Db 1195 IFLTTLS-NVHENGRIHQENIOEIEKEALIEKVVLPOVHEATGSKNFLKIDILITGR 1253
QY 513 QN-----NIPLLTLSEQSHLIPDGNLSHFYQGDWTFMKSDEGHSILA 560
Db 1254 QNLSLVEVHVPLQNTSINNSTVQIHM-----EHFKRR-----KDETNSEGLV 1301
QY 561 NMTPKNVVPHPERQSTLVANTLMNTYSDMQAVOSMINTIAHGAVLFGTWGSAVSNLFYA 620
Db 1302 NKTR-----MVKN-----YPSQKNITTORSKRALGOFRLSTQW---LKTINCS 1342
QY 621 HDSSGKPIDNMHHRSGLYLGISTHSLDHSFCLAGQLGKSSDSFTSTETTSYIATV 680
Db 1343 TQCIKQID--HSEKMKR--ITKSSLSDSSVTKSTQY--NSSDSHIVKTSAPPIDLK 1396
QY 681 QAOLATSLMKISAQACYNESIHLELTK 707
Db 1397 RSPFONKFSHVQA-----SSYIYDFKTK 1419

RESULT 20
PCT-US94-13200-6
Sequence 6, Application PC/TUS9413200
GENERAL INFORMATION:
APPLICANT: Emory University
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 1100 Peachtree Street, Suite 2800
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30309

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/13200
FILING DATE: 15-NOV-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Padst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: EMU106CIP(2)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-815-6508
TELEFAX: 404-815-6555

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2319 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Mus musculus
PUBLICATION INFORMATION:
AUTHORS: Elder, F.
AUTHORS: Lach, D.
AUTHORS: Galschier, J.
TITLE: Sequence of the Murine Factor VIII cDNA.
JOURNAL: Genomics
VOLUME: 16
PAGES: 374-379
DATE: 1993

RELEVANT RESIDUES IN SEQ ID NO: 6; FROM 1 TO 2319
PCT-US94-13200-6

Query Match 2.9%; Score 132.5; DB 5; Length 2319;
Best Local Similarity 20.2%; Pred. No. 0.02;
Matches 151; Conservative 109; Mismatches 300; Indels 187; Gaps 35;

QY 46 FECT-----MESIFPAL-----CAHASQDDPLVIGNSICWFYSKLIHTDPREALF 92
Db 775 FKDSITPKNOMEKIEPQFEIAELKQSVSVSDMLLQCSHP-TPHGIFLSDGQEAIV 833
QY 93 K-----EKQDLIQNFRFLSFTDCSSKESPSTIIQKNGQLSRNNGSFSFC 139
Db 834 EALHHDHSPNAIDSNBSPSVTQLR-----PESHSEKVIPTPOPGLDRKSKSL--- 883
QY 140 RNHAEGSGA1SADAESLQHNVLFT-----AEE-----NSSKNGCAI 178
Db 884 ETTEVYWKRLGLQVSLPSNLMTTILSDNLKATFEKTDSSGFPDMPIVASSKLSSTAF 943
QY 179 QAOTFSLRNVSPISFARNRADIINGCAICCSNLICSGVNPFLFPGNSATNGAICCI 238
Db 944 CKRAYSLVCSHVPNLNASENSD-----SNILDS---TLMYSQSLPRDNLSTEND 991
QY 239 LNTSEK--GSLSLACNQTLEFASNAKEKGATY---AKIMVLRYNCPVFINNSAKIGC 293
Db 992 RLREKRFHGIALTVDNTLEKQVSLMKTKNTKYNSTTEKLTSPSTIENSTQLOD 1051
QY 294 A1-----AIOGCSLILAGEGYLFEQNSQRTSDQGLVNAIYLEKDALIS--SLEARNG 347
Db 1052 AILKVNSEIQ--EVTALLINDGTLIGKNSF-----YLRNHLNLTNTTSTKKN 1095
QY 348 DILF--FDPIVQESSKESPLVSLQASVTSPTPATASPLVIQTSANRSVIFSSERLSE 404
Db 1096 DIFHRKDEDPIPODEENTIMPFSKMLFLS-----ESSNMFKKTKNGNSL-----NSE 1142
QY 405 EERTPDNLTSOLOPIELKSGRLVKRAVLASAP---SLSDQALLIMAGTSKLTSSD 461
Db 1143 QHSPKQVLYLM-----FKRYVKNQSLSEKKNVYQDGFKNIGLKIDMAFPNNMS 1194
QY 462 LKATLSIPLHS---LDTEKSV---THAPNLSIOKIFLSN-----SCDENFENVELLSKE 512
Db 1195 IFLTTLS-NVHENGRIHQENIOEIEKEALIEKVVLPOVHEATGSKNFLKIDILITGR 1253
QY 513 QN-----NIPLLTLSEQSHLIPDGNLSHFYQGDWTFMKSDEGHSILA 560
Db 1254 QNLSLVEVHVPLQNTSINNSTVQIHM-----EHFKRR-----KDETNSEGLV 1301
QY 561 NMTPKNVVPHPERQSTLVANTLMNTYSDMQAVOSMINTIAHGAVLFGTWGSAVSNLFYA 620
Db 1302 NKTR-----MVKN-----YPSQKNITTORSKRALGOFRLSTQW---LKTINCS 1342
QY 621 HDSSGKPIDNMHHRSGLYLGISTHSLDHSFCLAGQLGKSSDSFTSTETTSYIATV 680
Db 1343 TQCIKQID--HSEKMKR--ITKSSLSDSSVTKSTQY--NSSDSHIVKTSAPPIDLK 1396
QY 681 QAOLATSLMKISAQACYNESIHLELTK 707
Db 1397 RSPFONKFSHVQA-----SSYIYDFKTK 1419

RESULT 21
US-08-728-470-10
Sequence 10, Application US/08728470
Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Matlare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

AUTHORS: Gitschler, J.
 TITLE: Sequence of the Murine Factor VIII cDNA.
 Patent No. 5744446
 JOURNAL: Genomics
 VOLUME: 16
 PAGES: 374-379
 DATE: 1993
 RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 1 TO 2319
 US-08-474-503-6

Query Match 2.9%; Score 132.5; DB 1; Length 2319;
 Best Local Similarity 20.2%; Pred. No. 0.02;
 Matches 151; Conservative 109; Mismatches 300; Indels 187; Gaps 35;

46 FEDCT-----MESLFPL-----CAHAGDDPLVVLGNSYCWTFSKLHIDPKALF 92
 775 FKDSTIPKNDMEKIEPQEEIAEMLKVQSVSDMLMLQOSH-TPHGFLSDGDEAII 833
 93 K-----EKGDLISQNFRLSFTDCSSKSSPSIIHQKNGQLSLRNGSMSC 139
 DB 834 EAIHDDHSPNADISNDEGPKAVTOLR-----PESHSEKIVFPQGLQLRSMKSL--- 883
 QY 140 RNHAEGSGGAIADAFSLQINLYFT-----AFEE-----NSSKNGKGA1 178
 DB 884 ETTIEVWKKRLGLQVSSLPNSLMTTLLSDNLKATFEKTDSSGFPDPVHSSSLSTTAF 943
 QY 179 QAOFTSLSRNVSPISFARNADLNGCAICCSNLICSGNVMDLPFTGSAATNGCAICCID 238
 DB 944 GKRAYSLVGSVPLNASENSD-----SNLDS-----TLWYSQSLSPDNLSIEND 991
 QY 239 LNTSEK--GSLSLACNOETLFASNSAKEKGAII--AKIHVLYRNGPVSEFINNSAKIG 293
 DB 992 RLLEKRFHGIALLTKONTLFDKDVSLMKNTKYNHSTNKLHTSPJSIENSTJLQD 1051
 QY 294 AI-----AIOGGSLSLAGESVLPFONNSQRTSDGLVRAIYLEDAIIS-SLEARNG 347
 DB 1052 AILKVNSEIQ---EVTALLHDGTLGKNT-----YLRLHMLNRTTSTK 1095
 QY 348 DILF---FDPVQSSSKESPSPSSLOASVTPYPATASPLVIQTSANRSVTFESSERISE 404
 DB 1096 DIFIRKDEDPIDODEMTIMPFKMLFS-----ESSNFKKTNGNSL-----NSE 1142
 QY 405 EEKTPDNLTSOLOPILKSGRLVKDRAVLAP---SLSDPQALLIMEAGTSLKTTSSD 461
 DB 1143 OEHSPKQVLVLM-----FKKYKKNQSFLEKKNKVTVEODGFTKNGIKLDMAFPHNMS 1194
 QY 462 LKATLSLPLHS---LOTEKSV--THAPNLSTOKIFLSN---SGDENFLENWELLSKE 512
 DB 1195 IFLTTL-S-NVHENGRRHNOEKNOIEIEKVALIEEKVYLPVHEATGSKNFKLDLILIGTR 1253
 QY 513 QN-----NIPLLTSLKQSHLHPDGNLSSHFGVQGMWTFSMKQSDCHSLIA 560
 DB 1254 ONISLEYVHVPLONITSINNSTVQIHM-----EHFFKRR-----KDETNSEGLV 1301
 QY 561 NMTPEKNVYPPPEROSTVANTLMTVYSDMOAVQSMINTIAGAVLFGTWSAVSNLEFYA 620
 DB 1302 NKTRF-----NVKN-----YPSQKNITQORSKRALGQFRLSTQW---LKTINCS 1342
 QY 621 HDSSGKPLDMMHRSLSGLFGISTHSDHSCLAAGOLLKSSDSFTSTETSYIATV 680
 DB 1343 TQCIKQID--HSKEMKF--ITKSSLSDSVVIKSTQT--NSSDSHIVKTSAPPIDLK 1396
 QY 681 QAOLATSLMKISAQACYNESIHEIKTK 707
 DB 1397 RSPFONKFSHYQ-----SSYIYDKTK 1419

RESULT 18
 US-08-670-707A-6
 Sequence 6, Application US/08670707A
 Patent No. 5859204
 GENERAL INFORMATION:

APPLICANT: Lollar, John S.
 TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
 NUMBER OF SEQUENCES: 40
 CORRESPONDENCE ADDRESS:
 ADDRESSER: Greenlee, Winner and Sullivan, P.C.
 STREET: 5370 Manhattan Circle Suite 201
 CITY: Boulder
 STATE: Colorado
 COUNTRY: USA
 ZIP: 80303
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/670,707A
 FILING DATE: 26-JUN-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/US94/13200
 FILING DATE: 15-NOV-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/212,133
 FILING DATE: 11-MAR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/864,004
 FILING DATE: 07-APR-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Greenlee, Lorraine L.
 REGISTRATION NUMBER: 27,894
 REFERENCE/DOCKET NUMBER: 75-95F
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 303/499-8080
 TELEFAX: 303/499-8089
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2319 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: YES
 ANTI-SENSE: NO
 FRAGMENT TYPE: N-terminal
 ORIGINAL SOURCE:
 ORGANISM: Mus musculus
 PUBLICATION INFORMATION:
 AUTHORS: Elder, F.
 AUTHORS: Lakich, J.
 AUTHORS: Gitschler, J.
 TITLE: Sequence of the Murine Factor VIII cDNA
 Patent No. 5859204
 JOURNAL: Genomics
 VOLUME: 16
 PAGES: 374-379
 DATE: 1993
 RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 1 TO 2319
 US-08-670-707A-6

Query Match 2.9%; Score 132.5; DB 2; Length 2319;
 Best Local Similarity 20.2%; Pred. No. 0.02;
 Matches 151; Conservative 109; Mismatches 300; Indels 187; Gaps 35;

46 FEDCT-----MESLFPL-----CAHAGDDPLVVLGNSYCWTFSKLHIDPKALF 92
 775 FKDSTIPKNDMEKIEPQEEIAEMLKVQSVSDMLMLQOSH-TPHGFLSDGDEAII 833
 93 K-----EKGDLISQNFRLSFTDCSSKSSPSIIHQKNGQLSLRNGSMSC 139
 DB 834 EAIHDDHSPNADISNDEGPKAVTOLR-----PESHSEKIVFPQGLQLRSMKSL--- 883

Db 495 TFWNITANORIVYSSIN-LSNGSLTLMSEGRSGGVEINNDITTGDDTGANLTIYSGG 553

QY 627 PUDMMHRSICLYFGISTHSLDHSFCLAQOLL-----GKSS 664

Db 554 WVDHKKISLCAOGNINITAKODAFKESNOVITGOGTTTSGNOKGFRNNVSLNGTCS 613

QY 665 DSEITSTETSYATVQALATSL---MKISAQACYNESIHLELTKYRSPKESFGSMH- 720

Db 614 GLOFTTKRTKKAITNKFEGLTINISGVNISMWLPKNEGYD-KFKGRTY-----WNL 665

QY 721 ---SVAQSEVCASIPVNSGCLFSSFSIFSKLQCFSGQDCHPSSGELRSPASSFR 777

Db 666 TSLVSESGEF--NLITDSRSGSDAGLTTPYNLNGISFNKDD---TTFVNERAVNF- 718

QY 778 NISLPIGI 785

Db 719 DIKAPIGI 726

RESULT 15

US-08-169-927-2

Sequence 2, Application US/08169927

Patent No. 5783441

GENERAL INFORMATION:

APPLICANT: Carl, Mitchell

APPLICANT: Dobson, Michael E.

APPLICANT: Ching, Wei Mei

TITLE OF INVENTION: Gene and Protein Applicable to the

TITLE OF INVENTION: Preparation of Vaccines for Rickettsia prowazekii and

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Counsel, Naval Medical R & D Command

STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.

CITY: Bethesda

STATE: MD

COUNTRY: USA

ZIP: 20889-5606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/169,927

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/742,128

FILING DATE: 08/09/91

ATTORNEY/AGENT INFORMATION:

NAME: Spevack, A. David

REGISTRATION NUMBER: 24,743

REFERENCE/DOCKET NUMBER: 75,976

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 295-6759

TELEFAX: (301) 295-1022

INFORMATION FOR SEQ. ID NO. 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1612 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-169-927-2

Query Match 3.0%; Score 134.5; DB 1; Length 1612;

Best local similarity 19.9%; Pred. No. 0.0064;

Matches 166; Conservative 128; Mismatches 299; Indels 243; Gaps 41;

QY 130 LRNNGS---MSEFCRHHAECSGGAISADAFSLQHNLYLTAFEENSSKNGAGATQATFSL 185

Db 393 YKNGNTAGVITFMAN-----GALVAST-----DPNIANTINALEAGAV 435

QY 186 SRNVSPISFARNRADLNGAI--CCSNLJCSGVNVPLEFNGNSATNGCAICISDLNTSE 243

Db 436 V-ELSGIHIAELRLG-NGSIFKLADGTIVNGVQNAQNNMNLAAAGSI---QLDGA 489

QY 244 -----KGSISLACNDETFASNSAKEKGALIAKIMVLRYPGVSFFINSAKIGAI 296

Db 490 ITDIDGNGGVNMAIQHITL-ANDASK-----ILLALDG-ANIT--GANVGAIH 534

QY 297 IO--SGGSLSLIAGEGSLVFCNNNSORTSDOGLVRAIYLEADALISLEARNCDLTFDP 355

Db 535 FOANGGITKLNTNNINLVNFDLITDTKGTGYDASSLTNNQTLT---NGSI----- 584

QY 356 VOESSKESPLPSLQASVSPPTATASPLVIGTSANRSVIFSSERLSEBEKPDNLTSQ 415

Db 585 -----GTVANFTKTLAQ---LNGSSKTLINAGDVAINELVLENNOSVQ 625

QY 416 LOOPTEL--KSGRLVKDRAVLSPSLSDOPALLINEAGTSKTS---SDKLATLST 469

Db 626 LNHNTYLITKTIANAAGQIIVAADPLNTN---TTLADGTNLGSAENPLSTIHFAAKA 681

QY 470 PLHS-LDPEKSVITHAENLSIOLIFLSNGSDENFENVELLSEQNIFPLTISKBSH- 527

Db 682 NADSIILNVGKGVNLYANNITND---ANVSLIFRSG-----GTSIVGTYGGQGHK 731

QY 528 ---LHLPDGNLSSHFGYQGMWFSWKDDECHSL--TANNTPKNVYPPHROSTL--VANT 581

Db 732 LNNILUNGTTVK---FLGDTTFNGKPIEKGSLIQISNNYTTDHYEASANTGTLLEVNT 788

QY 582 LNMYSQMAVQSNININIAAGAYLESTNGSAV---SNLFYADSSGKIDNMHHISLG 637

Db 789 -----DPTVYLNKQAGV-FGVLMQVIISGPNIVF-----NEIGN-----YG 825

QY 638 YLFCISTHSLDHSFCLAQOLLCKSSDSFTSTETSYATVQALATSLMKISAQACV 697

Db 826 IVHGIANNSISFENASIGTSLPLPSGTPDLVLTIKSTVGNGTVDNFAPIVYVSGIDSMI 885

QY 698 NESIHELKTKYRSKSGFGSMHSAVSGEVASIPVNSGCLFSSFSIFSKLQCFSGT 757

Db 886 NNG--QIICDKKNITIALSLSDNSITV-----NANTLYSGIRTKNNQGVTL 931

QY 758 QDGFEESSEIRSRASFFNISLPIGITFEKKSOKR---TYFFLAVYO----- 806

Db 932 SCGPNNPGFI-----YGLGLENGSPKLQVFTTVDYNLISJIANVYTN 977

QY 807 -----DLK-----RDVES-----GPVYLKNA 823

Db 978 DDVTLTGAGTDFDQAKIILGVSNGNANRFVDSSTSDPSRMIVATQAKGTYITLGA 1037

QY 824 V-----SWDAPMANLDSRAVMFRLTNORALHRLQTLNVSCVLRGQSHSYSLDGT 874

Db 1038 LVSNIGSLDTPVASV-----RFTGN-----DSGAGLOGNIIYSGNIDFGT 1076

RESULT 16

US-08-212-133A-8

Sequence 8, Application US/08212133A

Patent No. 5663060

GENERAL INFORMATION:

APPLICANT: Lollar, John S.

APPLICANT: Runge, Marschall S.

TITLE OF INVENTION: Hybrid Human/Animal Factor VIII

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kilpatrick & Cody

STREET: 100 Peachtree Street

CITY: Atlanta

STATE: Georgia

COUNTRY: US

ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk


```

Db 126 NRVTS-----NOISQLKGLDSDNGQVFL-----INF-----NGITIGKDAII 162
QY 240 NTSEKGSLSLACNOETLFASN-----SAKEKGAIAKHA--MVLRYNGEVSFINNSAKTIG 293
Db 163 NTNGFTASTLDISNENIKARNFTEQTKDALKAEIVNHGLITVKGDSVNLIGKVKNEG 222
QY 294 AIAIAGSGSLIAGEGCVLFQNNNSQRTSDQGLVRNAIYL-----EKDAI-LSSLERANG 347
Db 223 VLSV-NGCSISLLAG-----OKTITSDIINPTITVISAAPENAVNLGDIAPKAG 271
QY 348 DILFFDPIVOESSKESPLPSLSQASVTSPTPATASPLVIQTSANRSVIFSSERLSEEEK 407
Db 272 NI-----NVRATIRNQGKLSADSV 291
QY 408 TPBNLTSQLQPIELKSGRLVKDR-----AVLSAPSLSDPPQALIMEAGTSLKTS 460
Db 292 SKD-----KSGNIVLSAKEGEAEIGVISAQNOQAKGKLMITGDKVTLTKGA 339
QY 461 DLKLA-----TLSIPL-HSLDTEKSVTHIAP-----485
Db 340 VIDLSGKGEGETYLGDEERGCKNGKIQIOLAKKTSLEKSTINVSKEKGRAIWMDIALI 399
QY 486 --NLSIOKI-----FLNSGDENFEYENVLLSKEO-----NNIPLLTSLKESHLHL 530
Db 400 DGNINAGSGSDIAKTGGFVETSGHDLFIKDAIYDAKEMWLDLDFNNVSINMETAGRS-----455
QY 531 PDGNLSSHFGYQGDWTFSMKDSDEGHSLIANWTPKNVVPAPEROSTLVANTLM-----N 584
Db 456 --NTSEDEYTG-----SGNSAS--TPKR-----NKEKTLTNTLTLESILKKG 494
QY 585 TYSMDQA-----VQSMINTIAHGAYLF--GTWGSAY--SNLFYAHDS-----SGK 626
Db 495 TEVNITANQRIYVNSSIN-LSNGSLTLMSEGRSGGVEINNDITGDTGRCANLTIYSG 553
QY 627 PIDNMHRSGLYLFGLISTHSLDHSFCLAGOLL-----GKSS 664
Db 554 WVDVHKNISLGAQGNINITAKODIAFEKGSQVYTGQGITSGNOKGFRFNNSLNGTGS 613
QY 665 DSIPTSETTSYATVQAQATLSL--MKISAQYNESIHELKTKYRSFSEKQSGMH- 720
Db 614 GLQPTKRTNKYAITNKEGFLNISGRVKNISMVLPKNESGYD-KFKGRTY-----MNL 665
QY 721 ---SVAVSGEFCASIPVNSGSLFSSFSIFSKLQFGSGVODGEEBSGELTSSASSFR 777
Db 666 TSLWVSESGE--NLTIDSKQSDSAGCTLPYNLNGISFNKD---TPFVVERNAKRVN- 718
QY 778 NISLPIGI 785
Db 719 DIKAPICI 726

```

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,832
FILING DATE: 16-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Belkstreaser, Jerry M
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-404
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-832-2

Query Match 3.0%; Score 134.5; DB 1; Length 1536;
Best Local Similarity 19.4%; Pred. No. 0.0058;
Matches 153; Conservative 123; Mismatches 253; Indels 259; Gaps 38;

QY 120 IHHKNCQLSLRNNGSMSPCRNHAEGSGAISADAFSLQHNLYFTAFEESSKNGNGAIQ 179
Db 76 VVH--GTATMOYDGNKTIIRNSVD--AIINMKQFINDENVOPLQEN--NNSAVF 125
QY 180 AQTFSLSRNVPSPISFAARRDLNGAICSNLJCSNVNPLFTGNSANGAICCSIDL 239
Db 126 NRVTS-----NOISQLKGLDSDNGQVFL-----INF-----NGITIGKDAII 162
QY 240 NTSEKGSLSLACNOETLFASN-----SAKEKGAIAKHA--MVLRYNGEVSFINNSAKTIG 293
Db 163 NTNGFTASTLDISNENIKARNFTEQTKDALKAEIVNHGLITVKGDSVNLIGKVKNEG 222
QY 294 AIAIAGSGSLIAGEGCVLFQNNNSQRTSDQGLVRNAIYL-----EKDAI-LSSLERANG 347
Db 223 VLSV-NGCSISLLAG-----OKTITSDIINPTITVISAAPENAVNLGDIAPKAG 271
QY 348 DILFFDPIVOESSKESPLPSLSQASVTSPTPATASPLVIQTSANRSVIFSSERLSEEEK 407
Db 272 NI-----NVRATIRNQGKLSADSV 291
QY 408 TPBNLTSQLQPIELKSGRLVKDR-----AVLSAPSLSDPPQALIMEAGTSLKTS 460
Db 292 SKD-----KSGNIVLSAKEGEAEIGVISAQNOQAKGKLMITGDKVTLTKGA 339
QY 461 DLKLA-----TLSIPL-HSLDTEKSVTHIAP-----485
Db 340 VIDLSGKGEGETYLGDEERGCKNGKIQIOLAKKTSLEKSTINVSKEKGRAIWMDIALI 399
QY 486 --NLSIOKI-----FLNSGDENFEYENVLLSKEO-----NNIPLLTSLKESHLHL 530
Db 400 DGNINAGSGSDIAKTGGFVETSGHDLFIKDAIYDAKEMWLDLDFNNVSINMETAGRS-----455
QY 531 PDGNLSSHFGYQGDWTFSMKDSDEGHSLIANWTPKNVVPAPEROSTLVANTLM-----N 584
Db 456 --NTSEDEYTG-----SGNSAS--TPKR-----NKEKTLTNTLTLESILKKG 494
QY 585 TYSMDQA-----VQSMINTIAHGAYLF--GTWGSAY--SNLFYAHDS-----SGK 626
Db 495 TEVNITANQRIYVNSSIN-LSNGSLTLMSEGRSGGVEINNDITGDTGRCANLTIYSG 553
QY 627 PIDNMHRSGLYLFGLISTHSLDHSFCLAGOLL-----GKSS 664

```

QY 563 TRKNVPHPEROST-----LVANTLNTYSMDQA 591
DB 677 TRFKYISSNSKGLTTOYRSSAGYNGVNGNMFNLKCAKYNFKLKPENMNT-SKPLP 735
QY 592 VQSMINTIHAHGYLFEGTWSAVSNLEFYAHDSGKPID-NMHRSLGYFEGISTHSLDDH 650
DB 736 IRLANITATG-----GSVFEDYANISGAGELKSEINISGANTFLNSHVRGDD 788
QY 651 SFCLAGOLLKSSDSFIITSTETTSYIAVQALATSLMKISAQACYNESJHELKTK-YR 709
DB 789 AFKI-----NKDLTINATNS-----NFSLRQTKDOFYD 816
QY 710 SFSEGEFGSMHVAVSG 726
DB 817 GYARNAINSTYNISILG 833
RESULT 7
US-08-617-697-9
Sequence 9, Application US/08617697
Patent No. 5977336
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5977336-typeable Haemophilus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STEET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstesser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1599 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-617-697-9
Query Match 3.2% Score 146 DB 2: Length 1599;
Best Local Similarity 18.5% Pred No. 0.00051;
Matches 153; Conservative 143; Mismatches 312; Indels 218; Gaps 39;
QY 73 NSYCMFVSKLHTIDPREALFKENG-----DISIONRFLSTDCSSKSSPSIHKONG 126
DB 143 NGCVFLINPGITIGKDAINTNGFTASTLDSINENIKARNFTLEDTKOKALAEI-VNHG 201

QY 127 QLSLRNNGSMSTCRNHAEGGCAISADAFSLQHNLTFTAFEESSKNGCAIQ---AQTF 183
DB 202 LITVCKDGSVNL-----CGKVK-----NEGIVSNGSISLLAGQKI 239
QY 184 SISRANSP-----ISPARRADLNGAICCSNLCSGNYNPLFFPGNSATNGALCCISD 238
DB 240 TISDIITNPTITTSIAPEDEA-INLGDIATK-----GGNIN-----VRA 277
QY 239 LMTSEKSLSLACNOETLFPASNSAKEGGAIV-AKHVILRYNGPVSPFNNSAK-----290
DB 278 ATRNKGRKLS-----ADSVSKDKSGNIYLSAKEGAELIGVYISAQNOQAKGKGLMI 328
QY 291 IGCAITAIQSGGSLSLIAGGCVLPFONNSORTSDQGLVRAITLLEDAILSLEARN-----346
DB 329 TGDKVTYLTGAVYIDLSGKEGETYLOGDER-----GCKKNGIOLAKKTTLEKSTINVSGK 384
QY 347 -----GDILFPDPIVQESSKSPSSLSQASVTSPTPATASPLVOTSA-----391
DB 385 EKGRALVWGDIALIDGNT-----NAQGDIAKTG-----FVETSGHYLSD 427
QY 392 NNSVIFSSERI,SEEEKTPDNLISQLOPIELKSGRLVLDRAVLASPSISODPOLIME 451
DB 428 DNATVTKEMLLD-----PENVTIAPSASRVELG-----ADRSNHADEVK-----VTLKK 474
QY 452 AGSLKTSDDLATLSTPLH--SLDTEKSVTHAPNLSIQK---TFLSNGDENFYENV 506
DB 475 NNTSLVTLTNTTISNLKSAHAVNITARRKLTIVNS-SIHERGSHLILHSEOGG--QGV 531
QY 507 EL---LSKQNNPIPLTLTSKEQSHLIPDG---NLSHFVGQDPTFGMKSDGCHSLI 559
DB 532 QIDKDTSEGNLTISGQWVDYHKNTITLGSGLNATTK---EGDIAFDKSGRNNLTIT 588
QY 560 ANNTPKNVYPHPEROSTLVANLTWN--TYSDM---QAVQSMINTIAHGAIVLEGTWS 612
DB 589 AUCITSGNSNGFRNNSVLSLIGKLSFTDSREDGRRTKGISMKFGDTINISGT---645
QY 613 AVSNLEFYAIDSSGKPIDNMHHSGLYLF-GISHSLDHSFCLAGOLLKSSDSITST 671
DB 646 -----VDISMKAPKVSFWRDCKRTYNNVTTLVN-----TSGSKRNLSTI 684
QY 672 EFTSYATVQALATSLMKISAQACYNESJHELKTKYRSFSGFSGMHSVAVSGEVCS 731
DB 685 DSTGSGST-----GPSIRNRELNGITFNKATFNIAOGSTANSTIKAS 726
QY 732 I-PIVNSGSGLEFSFISFKLOGFCGTODCFESSGEIRSFASSFRNISLPITGTFEKK 790
DB 727 IMFCKSN-----ANVALFNEIDISVS-----GGSVNFKLASSNSITOP-GVII--K 770
QY 791 SOKTRTYFFLAGYIODLKRDESGFVLLKNVSDADAMANDSR 836
DB 771 SQN-----FNVSGSTLNLKAEGSTETAFSITENDLNLATGNTIIR 812
RESULT 8
US-08-728-470-9
Sequence 9, Application US/08728470
Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STEET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

APPLICATION NUMBER: US/08/469,880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS-VG
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-469-880-4

Query Match 3.3%; Score 151.5; DB 2; Length 1477;
Best Local Similarity 18.9%; Pred. No. 0.00013;
Matches 151; Conservative 123; Mismatches 274; Indels 249; Gaps 36;

73 NSCYMFVSKLHTDPKEALFKKGG-----DLSTQNFRLSTFDCSSKSS--PSIIHQ-- 123
143 NCQVFLINNGITIGKDALINTNGFLASTLDSNEIKARNFTFEOTKALAEIVNHDL 202
124 ---KNGQSL-----RNGSMSFECRNHAEGSGAIS-----ADAFSLQHNLYFLTA 165
203 ITVGKGSVNLIGKAKVNEGVISV-----NGGSISLACGKITISDLINPIITYSIAA 255
166 FEENS-----SKG-----NGCAIQAFPSLSRNVSPIFANKHADI,NCGALC 207
256 PENEAVNLGDIKAGNINVRATIRNKGKLSDSVSKSGIVLSAKGEAEVIGVIS 315
208 CSNLICSGVNPFLFPGNSAT--NGCATCCISDINTSEKGLSLACNQFTLFAFSNAKE--K 265
316 AQNOQAKG--GKLMITGDKVTLKTGAVI---DLSCKEGG-----EVLQSGDEGECK 362
266 GGAIVAKHMYLRNGPVSEFINNSAKIGAIATQSGSLSLAGEGSLVLFQNNQSRTSDOG 325
363 NCQIQAQKTSLEKG--STINVSCKEKGGAIVWC--DIALIDGNIINAQSGSDIAKTG 416
326 LVNNA---IYLEKDALISLEARNCDILFDPPIVORSSSKESPLPSLSQASVSPPTATA 382
417 FVETSGHDLFIKDNATVDAKEW---LLDFDNY---SINAEPL----- 453
383 SPLVIOISANRSVIFSSERSEFEKTPDNLTSOLOPIELK-----SGRL----- 427
454 -----FNNTCINDEFPICTGASPKKNSKELKTLTJNTTISNYLKNAWTM 499
428 VLKRAVLASPSLSDPOLLIM---EAGTSKTSDDLKATLSTPLHS---LDTEKSV 480
500 ITASRKLTIVSSINIGNSHLILHSKQGRGQYQJODLITSCGNLTIYSCGVAVDHKNT 559
481 TIHAPNLISOKIFILNSGDNIFE---NVELLSK-----KONNIPL-----L 519
560 TLDOGFINTAASVAVEGGNNKARDAANAKIVAGVTYITTEGCKDRFANVSLNGTGKGL 619
520 TLSEKSHLHPDGNLSSHGQYQDWTFS-----WKSDSGHSLI-----ANW 562
620 NIISSVNL---THNLSCTINISGNITITNTTKRNTSYWQTSYSHDSHNVNVALNLFCGANF 676
563 TPKNYVPHPEROST-----LVANTLMNTYSDMA 591

Db 677 TFIKYSNSKGLTYORSSAGVNGVNGMSTNLEKAGKVNKRLKPNEMNT--SKPLP 735
QY 592 VQSMINTIAHGAVLFGWCSAVENLFYAHDSCKPID-NMHHRSGLYLFQISTHSLDH 650
Db 736 IRLPLNITATG-----GSVFIDYANHSGRGAELKMSFINISNGANFLNSHVAGDD 788
QY 651 SFCLAGOLGKSSDSFSTSTETTSYATVOALATSLMKISAOCINESHEIKTK-YR 709
Db 789 AFKI-----NKDLTINATNS-----NFSLRQTKDQFYD 816
QY 710 SFSKEGFGSHSVAVSG 726
Db 817 GYARNAINSTYNISILG 833

RESULT 5
US-08-728-470-4
Sequence 4, Application US/08728470
Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Matlare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-4

Query Match 3.3%; Score 151.5; DB 2; Length 1477;
Best Local Similarity 18.9%; Pred. No. 0.00013;
Matches 151; Conservative 123; Mismatches 274; Indels 249; Gaps 36;
73 NSCYMFVSKLHTDPKEALFKKGG-----DLSTQNFRLSTFDCSSKSS--PSIIHQ-- 123

```

Db 143 NCOVFLINPGITIGKDAIINTNGFTASTLDISNENIKARNFTEOTKDKALAEIVNHGL 202
QY 124 -----KNGQLSL-----RNNGSMFCRNHAESEGGGALS-----ADAFSLQHNLYFTA 165
Db 203 ITVKGDSVNLIGKVKYKNEGIVISV-----NGGSISLLAGOKITTSIDIIINPTITSIAA 255
QY 166 FEENS-----SKG-----NGCAIOAOTFSLSRNVPISFARRADLNGAIC 207
Db 256 PENEAVNLQDIFAKGNINVRATIRNOCKLSDSVSKDSKGNIVLSAKGEAEIGVIS 315
QY 208 CSNLICSGVNPLEFTGNSAT-NGAICISDLNLTSEKSLSLACNOETLFASNAKE-K 265
Db 316 AONOQAKG--GKLMITGDKVFLTKGAVI---DLSCKEGG-----ETVLGDERGEGK 362
QY 266 GGAIVAKHWLRYNGPVSPFINNSAKIGAIALIOGGSLSILAGESVLFQNNQSORTSDOG 325
Db 363 NGIOLAKKTSLEKG---STINVSCKEKGFAIYWG---DIALIDGINAOGSGDIKTTGG 416
QY 326 LVRNA---IYLEKDALISLEARNGDILFPDPIVOESSKESPLPSISQASVSPPTATA 382
Db 417 FVETSGHDLFIKNAIYDAKEW---LLDFDNV---SINMEDPL-----453
QY 383 SPLVIOTSANRSVIFSSERLSEEEKTPDNLTSQLQPIELK-----SGRL-----427
Db 454 -----FNNTGINDERPTGTGEASDPKKNSKELKTLTNTTISNYLKNAMTMN 499
QY 428 VLKDRAVLSAPSLQDPOALLIM---EAGTSLKTSDDLKATLSIPLHS---LDTEKSV 480
Db 500 ITASRKLTVNSINIGNSHLILHSKQGGGVQIDGDIITSKGNLTITSYGWVDVHKNI 559
QY 481 TIHAPNLSIOKIFLNSGDENFYE---NVELLSK-----EONNIPL-----L 519
Db 560 TLDGFLNITAAVAFEQGNKARDAANAKIVAOGTVTITGEGKDFRANNVSLNGTGKL 619
QY 520 TLSKQSHLILPDGNLSHFGYOGDWTF-----WKDSDEGHSLL-----ANW 562
Db 620 NIISSVNLT---TNHLSGTINISGNITITNOTTRKNITSYWTSHDSHMVNSALNLETGANF 676
QY 563 TPKNYPHPEROST-----LVANLTMTYSMDQA 591
Db 677 TFIKYSISNKGTLTOYRSSAGVNFNGVNGMNSFLKEGAKVNFKLKNENMNT-SKRLP 735
QY 592 VOSMINTIAHGAYLFGTWSAVSNLFYAHDSGKPID-NMHRSLQYLFGISHSLLDDH 650
Db 736 IRLANITATAG-----GSVFEDIYANHSRGAEELKMEINISNANFTLNSHVGGDD 788
QY 651 SFLCAAOQLLCKSDSRITSTETTSYIATVOAQLATSLMKSIAQACYNESIHELKTK-YR 709
Db 789 AFKI-----NKDLTINATNS-----NFSLRQTKDUDFYD 816
QY 710 SFSKEGFGSMHSVAVSG 726
Db 817 GYARNAINSTYINISILG 833

```

RESULT 2
 US-08-302-832-4
 Sequence 4 Application US/08302832
 Patent No. 5603938
 GENERAL INFORMATION:
 APPLICANT: Barenkamp, Stephen J
 TITLE OF INVENTION: High Molecular Weight Surface Proteins
 TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Shoemaker and Matarate, Ltd.
 STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 CITY: Arlington
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22202-0286
 COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302.832
FILING DATE: 16-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US pct/us93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Becklesser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-404
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0810
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-302-832-4

Query Match 3.3%; Score 151.5; DB 1; Length 1477;
Best Local Similarity 18.9%; Pred. No. 0.00013;
Matches 151; Conservative 123; Mismatches 274; Indels 249; Gaps 36;

QY 73 NSYCWPFYKLIITDPKALFEKKG-----DLSTONFRFLSFDSCSKRESS-PSIIHQ-- 123
Db 143 NCOVFLINPGITIGKDAIINTNGFTASTLDISNENIKARNFTEOTKDKALAEIVNHGL 202
QY 124 -----KNGQLSL-----RNNGSMFCRNHAESEGGGALS-----ADAFSLQHNLYFTA 165
Db 203 ITVKGDSVNLIGKVKYKNEGIVISV-----NGGSISLLAGOKITTSIDIIINPTITSIAA 255
QY 166 FEENS-----SKG-----NGCAIOAOTFSLSRNVPISFARRADLNGAIC 207
Db 256 PENEAVNLQDIFAKGNINVRATIRNOCKLSDSVSKDSKGNIVLSAKGEAEIGVIS 315
QY 208 CSNLICSGVNPLEFTGNSAT-NGAICISDLNLTSEKSLSLACNOETLFASNAKE-K 265
Db 316 AONOQAKG--GKLMITGDKVFLTKGAVI---DLSCKEGG-----ETVLGDERGEGK 362
QY 266 GGAIVAKHWLRYNGPVSPFINNSAKIGAIALIOGGSLSILAGESVLFQNNQSORTSDOG 325
Db 363 NGIOLAKKTSLEKG---STINVSCKEKGFAIYWG---DIALIDGINAOGSGDIKTTGG 416
QY 326 LVRNA---IYLEKDALISLEARNGDILFPDPIVOESSKESPLPSISQASVSPPTATA 382
Db 417 FVETSGHDLFIKNAIYDAKEW---LLDFDNV---SINMEDPL-----453
QY 383 SPLVIOTSANRSVIFSSERLSEEEKTPDNLTSQLQPIELK-----SGRL-----427
Db 454 -----FNNTGINDERPTGTGEASDPKKNSKELKTLTNTTISNYLKNAMTMN 499
QY 428 VLKDRAVLSAPSLQDPOALLIM---EAGTSLKTSDDLKATLSIPLHS---LDTEKSV 480
Db 500 ITASRKLTVNSINIGNSHLILHSKQGGGVQIDGDIITSKGNLTITSYGWVDVHKNI 559
QY 481 TIHAPNLSIOKIFLNSGDENFYE---NVELLSK-----EONNIPL-----L 519
Db 560 TLDGFLNITAAVAFEQGNKARDAANAKIVAOGTVTITGEGKDFRANNVSLNGTGKL 619
QY 520 TLSKQSHLILPDGNLSHFGYOGDWTF-----WKDSDEGHSLL-----ANW 562
Db 620 NIISSVNLT---TNHLSGTINISGNITITNOTTRKNITSYWTSHDSHMVNSALNLETGANF 676

```


Db 329 ICAYVNLQGRPLEFSNRKAGCALFICHSVDDPAKQSTLTLLSSEGILAPQGMNLN 388
321 TSDGLVNAIYLEKDALLSLEARNGD-ILFPDPIOESSKSEPLPSLSQAVSTSP 379
Db 389 TKPG--IKNATTEVAGGIYLSAQGSRLYFYRPIH-----SLPTTSP 431
380 AYASPLVLOTANSRNVIFSSERLSEEP-KIPDNLTSQLOQPIELKSGRLYLKRAVLAP 438
432 SNKDTITANAGSAGSVYTSKGLSTELLIPANTTILLGLTIVASGLKITDAAVNVL 491
Qy 439 SLSDPPQALLMEACTSLKTSDDLKATLSIPLASLDTKSVTHAARPLSIQKIFLS--- 495
Db 492 GFATQSGQLTLGSG-----GTLGLATPTCAPAAVDTICKLAFDPFSLKRFVSAV 545
Qy 496 NSGDFENFENNELLSKEONNIPILTLSEOSHLL-----PDGMLS--SH 538
Db 546 NAGTKNVTLLTALVLDHDTDLTDWVSLQSPVALPIAFVFGATVTYKIGFPGELIAPSH 605
Qy 539 PCYQGDWTFESMK-----DSDEGSHLIAMWP-----KNVYHPRKOSTLV 578
Db 606 YGQGMKSYTSRPLLPAPDGFPGGSPSPANLAVMNSDTLVRSYILDEPERGEIV 665
Qy 579 ANTLNNTVSDMOAVOSMIN---TIAHGATVLFKMGSAVSNLFPYAHDSGKPIDNMHHS 635
666 SNSIMISFLQNAESDILLQDVLIDHPLSI---TAKAGAVENTPRGCHGFGSGRY 720
Qy 636 LGYLEGISTHSDHSPCLAGOLLGK-SSDSFTSTETTSYIAVQALTSIMK---- 690
Db 721 GGYOALISMNTDHTTTLTGLSFGOLKCKTNANPYRSCQWYLSFHCQFPIYVQKSEAL 780
Qy 691 ISAQACYNESIHELKTKYKFSFK--ECFGSMHSAVAGCWCASPIVSNQGLSPFSIF 748
Db 781 ISMKAAVYSKNHLNTTYLPRDKAPKSGQGMHNSYVLLIAEHP-----FLNMCLL 832
Qy 749 SK-----LQGF-----SCTQGFESSCEIISFSAASSRNIISLPIGITFE----KK 790
Db 832 TRPLAQAAMDLSGFISAEFLGQMSKFTETCDLOKSFSGKQYVNSLPICSSQWFTPPKK 892
Qy 791 SOKRTTYVFLGATYQDLKRDVESGPVLLKNVSWDAPMANL----- 833
Db 893 APSTLTITKL--AKKPDYIVPNPNJYTVVSNQESTISIGCANLRHGLFVQIHVVYDITE 949
Qy 834 DSRA---YMFRLTNORALHRLQT 853
Db 950 DTQAFNLTFQCKNCFTHNRVST 972

RESULT 25
H72074
polymorphic membrane protein E/F family CP0286 [imported] - Chlamydotheca pneumoniae (strain C) [NCBI:taxon:100881]
C:Species: Chlamydotheca pneumoniae, Chlamydotheca pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: H72074; E81593
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lamme, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nelson, M.; Gwyn, M.; Nelson, M.; Deboy, R.; Kolonay, J.; McElarthy, G.; Salzberg, S.L. Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydotheca pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: H72074
A:Molecule type: DNA
A:Residues: 1-938 <AR>
A:Cross-references: GB:AE001631; GB:AE001363; NID:g4376750; PIDN:AD18608.1; PID:g4376750
A:Experimental source: strain CML029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, J.; Dodson, R.; Gwyn, M.; Nelson, M.; Deboy, R.; Kolonay, J.; McElarthy, G.; Salzberg, S.L. Nature Genet. 21, 385-389, 1999
A:Title: Genome sequences of Chlamydotheca trachomatis Mopn and Chlamydotheca pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: E81593
A:Molecule type: DNA
A:Residues: 1-938 <RE>
A:Cross-references: GB:AE002190; GB:AE002161; NID:g7189209; PIDN:AAE38143.1; PID:g7189209
A:Experimental source: strain AR39, HL cells

C;Genetics:
A;Gene: pmp_15; CP0286

Query Match	11.9%;	Score 540.5;	DB 2;	Length 938;
Best Local Similarity	25.2%;	Pred. No. 3.9e-26;		
Matches 229;	Conservative 122;	Mismatches 357;	Indels 201;	Gaps 38

```

0Y 124 KNGOLSLRNKNSMFCRRNHAEGSGALISAD---AFSLQHNYLT----- 164
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 77 KNLLIS-ENIGONVEFEKRVKVCNCGGAIYAQNCTISKNQYAFETTLVSDNPTATGSDL 135
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 165 -----AFEPNSKONGAIOAOV-FISLRNVSPISPARNA-----DL 201
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 136 GGALFALNCSTTNMLGOGTEVDNLALNKGALLTETMLSTKDKKGPLIYIKONRALMSDSL 195
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 202 NGAICCSNLICSGNVNPLPFTGNSATNGAICISDLINTSEKGSL----- 247
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 196 GGGIYSGNSLMEBNSGAIQIUTSSSSGCGIGFTQTLITSSKKLIEISENSAFANNYG 255
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 248 -----SLACNOE-TLFASNSAKEKGAIYKKNVLYRYNGPYVSTINNSAKI 291
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 256 SNFNNGGGGLTTECTTILNNEGVLFPNNNOSOSNGAIHAKSIIEKENGVPVLTATR 315
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 292 GGAI-----AIGSGSLIAEGSVLFRONNSORTSDGL-----VRNAIYEKDAIISLEA 344
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 316 GGALLNLSAGSGNSFSLISADNGITINNNT--ASKIALPPIRRNAIHSIPNNL-QIGA 372
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 345 RNG-DLIFEDPIYVESSESKEPSLPSLOASVSTPPTAASPLY--IUTSANRVSFSSER 401
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 373 RRCVYVLFYDPI-----EHELPSF-----PILEFNETGHTGVLPSEGH 412
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 402 LSEEKTPDNLTSOLOPIELKSRLVLYKDBAVALSAPSLSODPOALL-----IMEAGT- 454
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 413 VNONTEDEMNFEYSYLRNTESLROGVULVLEDGAGLACYKFFORGGLTLLGOGAITYTAGTI 472
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 455 -----SLKRS--SDIKIATLSIPLSL-----DTEKSVIHAAP 485
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 473 PTPSTPTTVGSTITLNIATIDLSIISFOAOKIWTYPTKGSTYTEDSNPTIITSG- 531
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 486 NLSIOKIFLNSGDENFENVELLSKEONNIPILT-----SKEOSHILPDGNLSHR 539
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 532 -----TLLRNSNNDPYDLSLD-LSHLEKVPRLIYIDVAOKINSQDLSLNGEHN 585
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 540 GYGDWTFSMKD-----SDEGHSLT-ANMTPKIVVPIREFQSTLVANTLYNTY 586
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 586 GYGCIWSTWYVETTTITNPITSLGANTKHKLYLAWMSPLGRPERGGEFTINMAL-- 642
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 587 SMMQAVQSMIMVIAHGAYULFGT-----GSAVS-----NLFYANDSSGKRLDMNNHRSIG 637
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 643 -----QSAVYALLA--GHSLISSWDEKGNHAASLOGIGLVLHOKDKNG--FKGRSHMTG 652
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 638 YLFGLSTSLDHSFCLAGOLLKSSP-SPTSTETTSYIA--TYVQOLATSLMKISAQ 694
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 693 YSATIEALNISQSPNSLQFAFPFKAAHEBQONTSIHNFQSGMCINTLFEKIRLSVS 752
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 695 ACYNESIHLELTKTRRSK-EGFSWHSVAASGEV-CASIPYVNSGGLFSSFSISFKLO 752
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 753 LAYMTSEHTITWOGLEEGNSGSPFNHTLAGALSCVFLP-OFHCESL--QIYPEFTAL 809
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 753 GHSQPODFEFSGEBIRFSA-SSPRNLSIPIGITFE-KKSQKPTUYVFLGAVIDOLK 810
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 810 AIRGVLAAFOESGDHAREFSJLHRLTIVSLVGGIRASMKNNHRRPLWLTEISRSTLYR 869
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 811 DYESGPVVLKRNAAWMDAPMNLDSRAVMEFRGLTIQORALNRLTOL--LNVSCVLRGQSHYS 869
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 870 QDPELHAKSLILISOGWWTQATPVITYNALGIKVKNTMQVFPKVTLSLDYASDISSTLSHY 929
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 870 IDLGTYTF 878
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 930 INVASMRF 938
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

QY 376 SEPPTATASPLVQTSTANRSVIFSSSEFLSEETKTPDNLTSOLOPILKSGRLVKDRAVL 435
DB 419 -----GILFESSTYIPDTSTSRDDPISHFRNHIGLYNGTLALEDRAEW 461
QY 436 SAPLSQDOPALLI-----MEACTSLKTSDDLKATLSLPLHSL----- 474
DB 462 KYKKEQFGGCTLRGSRVFEFTTDEOSSSSVGVININNLAINLPISIGNRAPKLMIR 521
QY 475 -----DREKSVYTHAPNLSTLQKIFLNSGDENF--YENVLLSKRONNIPLL-TL 521
DB 522 PGSSAPRYSEDNNPIINLSGP-----LSLLDENLDPDYAD-LAQP IAEVPLLYLL 572
QY 522 SKEOSHLH-----LPDG-NLSSHEGYGDMTFSSKD-----SDEG-----HSLIANWT 563
DB 573 DVYAKHINTDNFPEGLNTTQHYGYGVMSPVYIEFTITTSDEEDVNTLHQLGDMWT 632
QY 564 PKRYVHPEROSTLVANTLMTNTYSDM-----OAVOSMITIAHGCAYLPGTGSVAVSNL 617
DB 633 PGYKYNPEKKGDIASAFNOSHNLFAILRYTOGQGIAPTPASGA-----TR 661
QY 618 FYAHDSGKPIDMHHKSLQYLFPCISPHSLDHSFCLLAQGLLKSSDSFTITSEPTTSYI 677
DB 682 LAFVHONSNDKGFHEATQYSLGTTSTNTASNSFCVNFSQLFSNLYESH-SUNSVASHIT 740
QY 678 ATYQAOLATLSMK-----ISAOACYNESIHELKTKYRSFSKEGSGMISVAVSGEVCASIP 733
DB 741 TTVVALDITNPMLOERFSTASIAVSNNHIAKASYSGKIQTECKCYSTTLGAALSCSIS 800
QY 734 IVNSGQIFSSSFISFKLOQFSGTODCFESSGEIRSFSA-SSFRNLSLIGITFEKSKQ 792
DB 801 LQMRSRPL--HFTPTQIAIVRSNOTAFODESGKARKFSYHKPLNLTVLVGLQISAMEK 858
QY 793 -KRTTYVYFGLAYTDLKRDVESGPVLLKNAVSDAPMANLDSRAYMFRLTNORALH-R 850
DB 859 FRLLPTVNIETLAVQPVLYLOONPEINVSLESSGSMLLSTGLARNALIAFGRNDIFLPPK 918
QY 851 LQTLNVCYLRCQSHSYSLDGLTTFRF 878
DB 919 LSVFLDYQGSVSSSTPTTHYLACTFKF 946

RESULT 22

D72067
polymorphic membrane protein A family CP0213 [imported] - Chlamydia pneumoniae (stra
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: D72067 DB1601
R:Kilman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: D72067
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-947 <ARN>
A:Cross-references: GB:AE001638; GB:AE001363; NID:94376819; PIDN:AAD18679.1; PID:9437682
A:Experimental source: strain CML029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mogen and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150253
A:Accession: D81601
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-947 <REA>
A:Cross-references: GB:AE002182; GB:AE002161; NID:97189140; PIDN:AAF38083.1; PID:9718914
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: pmp_19; CP0213

Query Match 12.5% Score 566; DB 2; Length 947;

Best Local Similarity 25.3%; Pred. No. 9.6e-28;
Matches 252; Conservative 122; Mismatches 376; Indels 246; Gaps 41;

QY 54 LPPALCANASODDPLYVIGNS-----YCWFSKLIHTDPKALFKKGDLS 99
DB 27 LPLSGISHGDELELFTLRSSSPTKTYSLRKDFVICDPAGN-SIHKPGAAFLNLKGLDF 85
QY 100 ION-----FRLSFTCCSSKESPSTIHKQNGLSR 131
DB 86 FINSPLAALTKNHLGARGAGLFSESNTFKGLSHLVLENNEVGVL-TTSGLSFI 144
QY 132 NNGSMFCRNHAGSGAL-----SADAFSLQHN--YLE-----TAFENSSKNGAGAT 179
DB 145 NNTSVLCNNNISTGPGGALLGGRKSKALFPFRNKGITLFLANKAVNDESPRGYGVANS 204
QY 180 AOTFSLSRVNSPISFA-----RNRADLNG-----GAICCSNLICGNNVPLEFTGN 225
DB 205 ----SISPG-SPITFADNGELLFQENEGELGAIYNDQGAIFEN-----NFQTSFSSN 254
QY 226 SATNGAICISDLNTEKGSLSLACN-----QETFAANSKAKGAIYAKHMYR---- 277
DB 255 KASFGAV-----YSRYCNLYSOWGDVLTFTNAAKVGGAHADYVHLRDK 301
QY 278 -----YNGPVFTNSGA--KIGGAIAIGSGS-LSILAGE 309
DB 302 GSTVFEENSTAGATAYANVACDINNAGPYRFINNSALGNGAIYMQATGILRLHNG 361
QY 310 GSVALPQNNQORT-----SDQGLVNAIYLEKDALISLEARNGD-ILFDPDIYDESSS 361
DB 362 GDIEFGCNKRVQGFHSHINSTNFTNNAITIGAPREFSLANEGHRICFYDPIISATEN 421
QY 362 KESPLSSLSQAVSTSPPTATASPLVIO-----TSANRSVIFSSRLSEETKTPD-NLTSOL 416
DB 422 YNS-----LYINHORLLEAGAVLFGARLSPEHKENKNTSII 461
QY 417 QOPIELSGRLVKDRAVLSPSLSDOPQALLMEAGTSLEKT-----SSDLKATLSIP 470
DB 462 NQYRLLCGVLSIEGAIILAVRSFYQE-GGLLALGSGSKITTGQKNSKKIYITNLFEN 520
QY 471 LHSIDTEKSVTHAPNLSTLQKIFLNSG-----DENFYENVLLSKRONNIPLLTSLK 523
DB 521 LENLSSDPAEIRA-----TEKASIEISGVPRVYGHEFSFEYENHEYSKPYTTSIIISAKK 576
QY 524 EOSHLHPDGNLSS-----HFGYQGDWTFSW--KDSDEGHSLANMTPK-NYVPHR 571
DB 577 LVYAPSRHEKDIONLILAESEYMGQYQGSMEFSWSPNDTKKKTIIASVTPGEFSLOP 636
QY 572 EROSTIVANTLMTNTYSDMQAVOSMINTIAGGAYLFGT-----MGSAY----- 614
DB 637 KRKGSFIPTLMTSTFSLNIASNTVN-----NNTLNSEVITPLDHLCEVPGPYQIMEON 691
QY 615 -----SNLFYAHDSGKPIDMHHKSLQYLFGLISTHSLDHSFCLAGGLLKSSDSFT 669
DB 692 PROSSNNLIVQHAG-----HNVGARIPFSFT-----ILSAALTQLFSSSSQONVA 737
QY 670 -STTTTITATVQAOLATSLMKISAOACYNESIHELK--TKYRSFSKEGSGMISVAVSG 726
DB 738 DKSAQIILIGVSLNKKWQALSLRSSFSYEDSOVMKHNVPYKQTSR--GSMWNVWSG 794
QY 727 EVCASIPVNSGSL-FSSSFISFKLOQFSGTODCFESSGEIRSFSSFRNLSLIGI 785
DB 795 SVGSS---YAPKRIKRLKMPFDLDQYTKLVQNPFEVETGDPRIYESSSENTNLSLPGI 851
QY 786 TFEKK--SOKRTTYVYFGLAYIOLKRDVESGPVLLKNAVSDAPMANLDSRAYMFRLT 843
DB 852 ALEMRFTGSRSLFLQVSTSYIKDLRVNPPQSSASVALLNHTWIDQVPLGKALNITLN 911
QY 844 NORLHLRQTLNVCYLRCQSHSYSLDGLTTFRF 878
DB 912 STIKYKIVTAYMGISTSTORSSNLSANNAHGLSLSF 947

RESULT 23

```

Db      182 SVSFYONAAATFGGAIHSSGPLOIAVNOAEIRFQONTAKNGSGALSDSDIDIDQAAVYL 241
      221 FFGNSAT----NGAICCSIDLTSEKCSL--SLACNOETLFASNAKKGCAIVAKHHV 275
      242 FRENELTTAIGKGAIVCCLPITSGSSTPVPYTFSDNKQIVERNISIMGGATIAKKLS 301
      276 LRYNGPVSEIN-----NSAKIGCAIATGSGSLILAGESVLFQNNNSORTSDQGLVRNA 330
      302 ISSGCTLFINNITSYVNSQNLGAIADTGGELSLSAEKGTIFQGN--KTLSPFL--NG 357
      331 IYLEKDAIISLEARNG-DLFFDPPIVQESSKESPLPSLOASVTSPPPATASPLVIQT 369
      358 ILLONAKFLKLOARNGSYLIEFPITSEAD-----GSTQLNIDPK-----NK 402
      390 SANRSVIFSSERLSEEEKTPDNLTSOLOPTELKSGRIYKDAVLSAPSLSDPQALL 449
      403 EYTGITLFGESKSLAND--PRDKSTIPQNVNLSAGYLVKKEGAEVTSKFTSPSHLV 460
      450 MEAGSLKTS--SDLKATLSLPLSLDTEKSVTHAPNLSIQKIFLNSGD-----ENFY 503
      461 LDGTLALSKEDIALITGLAIDTSLSSSTAAVTKANTANKQISVDTIELSPGNAY 520
      504 ENVELLSKEONNIPLLTSLK--EQSHLHPDGN--LSSHEGYGDWTFSKKSDSDE--GIIS 557
      521 EDLRM--RNSOTFPLTLEFGAGSVTATGDFLPVSPHYFGQNMKLAWTGTGNKVGEE 578
      558 LIAMTPKNVPPPEROSTLVANTLMTYSDMAVOSMITIAHGAYLFCITMGSAVSNL 617
      579 F---WKINKPRPEKGNLVPNILMGNAVDRSLMVOETHASSLOTDRGLIDGIGNF 635
      618 FYAHDSGKPIDMMHRSILYLFGISLSDHSFLCAQAQLLGSKSDSFTSTETTSYI 677
      636 F--HVSASEINIRYRHNSSCYVLSVNNETIPKHYTSMAFSQLSRQDVAVSNNEYMYL 693
      678 ATVOAOLATSLMKISAQA-----CYNESIHELKTYRS 710
      694 GSYLYQYTTSLAGNIFRYASRNPVNVNGILSRFLQNLMTFHLCAIGHATNMKTDYAN 753
      711 FSKRGOSMHSVAVSGVCASIPVSMGSG--LFSSEFLSKLOCFSTOGDPESSSGEIR 769
      754 FPNVK--NSMNNNCWALECGSMPLVFNENGRLOGAIPFKLQLVAYQDQFKETIADGR 812
      770 SFASSEFRNLSLIGITFEFKSOKTRTYFFLAGAYIIDLKRDVESGPNVLKKNVADAP 829
      813 RFSNGSLTISVPLGIRFEXLALXODVLDYDFSPSYIPDIFRKDPSCAALVSDSDMLVP 872
      830 MANDSRAVYFRLTNQALHRLQTL--NVSCVLRGQSHSYSLDGTTFYRF 878
      873 AAHVSRHAFVSGSGTGRYHFNQYTELLCRGSIETC--RPHANNYNINCGSKFERF 922

RESULT 19
POLYOMPHIC membrane protein G family CP0761 (Imported) - Chlamydia pneumoniae (strain
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: E72130; G81541
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; WUID:99206606
A:Accession: E72130
A:Molecule type: DNA
A:Residues: 1-841 <RAN>
A:Cross-references: GB:AE001586; GB:AE001363; NID:94376263; PIDN:AND18172.1; PID:9437627
A:Experimental source: strain CWL029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Morn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; WUID:20130255
A:Accession: G81541

```

```

A:Molecule type: DNA
A:Residues: 1-841 <RAN>
A:Cross-references: GB:AE002235; GB:AE002161; NID:97189672; PIDN:AAF38561.1; PID:9718
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: pmp_2; CP0761

```

Query Match 15.6%; Score 708.5; DB 2; Length 841;

Best local similarity 28.5%; Pred. No. 7.8e-37;

Matches 241; Conservative 137; Mismatches 348; Indels 121; Gaps 31;

```

      107 SFTDCSSKSS-----PSIIHQ-----KNQQLSRNNGSNSEFCRNHAEGSGAISAD 153
      41 STFSFSSKTSASADGTNNYVFKDSVLENPVKTEGT-----STSCPKN-----DAAGD 89
      154 AFSLOHNYLFTAEENSSKNGCAIQAO-----FSLSRVNSPISPARNAD--LNG-GAI 206
      90 LNFGLGGFSFTSINIDATTAAGAISSSEANKVTLIS--GFSAISLTKSPSTVNTGLAI 148
      207 -CCSNLISGVNPLPFTGNATNGAICISDNLNTEKSGSLSLACNOETLFASNAKKE 265
      149 NVKGNLSLIDNOKVLLQDNFSTGDGAINC-----AGSLKIANNKSLIFGNSSSTR 200
      266 GGAIVAKHMYLRYNGPVSFINSAKI---GGAIAIQSGSLILAGESVLFQNNNSORT 321
      201 GCAIHITKNLTLSSCGEITFOGNAPTAAAGKGAIALADSCOTLISGDSGDIIFEGNT--1 258
      322 SDQGLV--RNAIYLEKALISLSEARNG-DLFFDPPI-VQESSKESPLPSLOASVTSPT 378
      259 GATGVSHSAIDGTSAKITLRAAGHTTYEFDPIYTGISTVADAL-----NINSP- 311
      379 PATASPLVIQTSANR---SVIFSSERLSE-EKKTDPDNLTSOLOPTELKSGRIYKDA 433
      312 -----DTGNKKEVTCIIVFSGEKLTEAKDEKNRPSKLLQNVAFNGVYLVKQDV 362
      434 VLSAPSLSDPQALLINEAGTSLKTSQD--LKLATLSLPLSLDTEKSVTHA-----PNLS 488
      363 VLSANGFSQDANKRLINDLQTSLVANTESIELTNLEINIDSLRNGKKIKLSATAQKDIR 422
      489 IOKIFLNSGDEMPEFVENVELLSKEONNIPLLTSLKEQSHLHPDGN--LSSHPYQGD 544
      423 IDRPVYLAISDESFTYQ--GFLENDHSTGILELDAGKDVIASDSRSIDAVQSPYGTQK 481
      545 WPFWSKDSDEGSHLIAMNTPKNVPPPEROSTLVANTLMTYSDMAVOSMITIAHGCA 604
      482 WTIINWSTQDK--KATVSMAGQSFNPALQEAFLVPLNLMGSLFDIVRSFQNFIELGTEGAP 539
      605 YLEGTWGSVAVSNLFEYAHDSGKPIDMMHRSILYLFGISLSDHSFLCAQAQLLGSKSS 664
      540 YERFVWAGISNVL--HRSGRNQKRFHVSQGAUVGASTRMPGGDTLSLGFALFARDK 597
      665 DSPITSTETTSYIATVOAQLATSLMKISA-----QACYN 698
      598 DYEMNTNFAKTYAGSLRLQDASLYSVSILLEGGLREILLPYVSKTLPDCSYGQLSYG 657
      699 ESIMELKTY-----RFSKGEFOSMHSVAVSGVCASIPV--SNGSGLFSSSIFSKLQ 752
      658 HTDHRMKTESLPEPPPLSLD--HTSMGCVYAGAGELGRVAVENTSGGTFQDETTPFYKVQ 716
      753 GFSCTQDGFEESSGEIRSFASSSFRNLSLPIGITEFKSOKTRTYFFLAGAYIIDLKRDV 812
      717 AVYARODSFVELCAISRDFSDSHLYNLAIPLGILEKRF--AEQYHVAVMYSDDVCRSN 774
      813 ESGPNVLKKNVADAPMANIDSRAYMFRLLTNQALHRLQTL--NVSCVLRGQSHSYSLD 871
      775 PKCTTLLSNQSGWKTKGSNLRARQATVQASGFRSLGAAALDELGNFEMRGSRSSTYND 834
      872 LGTTFYRF 878
      835 AGSKIKF 841

```

polymorphic membrane protein G family TC0263 (imported) - Chlamydia muridarum (strain N3)
C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: H81722
R:Read: T.D.: Brunham, R.C.; Shep, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Chinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbeg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae A839.
A:Reference number: A81500; MUID:20150255
A:Accession: H81722
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-987 <1ET>
A:Cross-references: GB:AE002293; GB:AE002160; NID:g7190298; PIDN:AAF39132.1; PID:g719030
A:Experimental source: strain N19g (Mopn)
C:Genetics:
A:Gene: TC0263

Query Match 16.7%; Score 758; DB 2; Length 987;
Best Local Similarity 26.7%; Pred. No. 7.5e-40;
Matches 246; Conservative 125; Mismatches 353; Indels 198; Gaps 28;

100 IONPRFSLFDSCSSKSSPSLIHQKNG-----QLSLRNGSMNSFCRHAH 144
121 IEAFDELSTLNCN---SLVSVPQTGGTTTVPNSCTIVSRDVLVKDIKKVSFYSMLVS 177
145 GSGGAIADAFSLQHNVLTFAPFENSCKGNGAIO-AOTFSLRNVSPISFARNRADLNG 203
178 GUGCAIDAGSLMNGIEKICTFOENVAQSDGACQVTKTFSAVCKKPIPSFLGNVACNKG 237
204 CAICCC-----SNLSCGVNPLPFTONSAT-- 228
238 GGAVALVKGAGGATDLSVFNANNTAVEFEENARSIRIGGIYSDGNIS--FLGNATKVF 294
229 -----NGAICICISDLNINSEKGSLSLACNOETLPASNS 261
295 LSNVAPSIYVDPAAAGCPADKDNVCDGALFCKNDTNI--GVVSFKUCGVVFESKN1 351
262 AKEKGAIYAKHVLRYRNGVSVFINSKATGCAIAIOSGSLSLGEGSVLPFONNSQRT 321
352 AAGKGAITVAKKLTISDCGVYOFGLGNVANDGAIYLVDOELPISAKKDIIFDGNLKR 411
322 SDGC-----LVRNAIYLEKDAIISLEAKNG-DLFFDPPIVOESSKESPLPSLQA 372
412 ATGGAATVHDVMAVSNAINSMATGQITTLRAKEGRILFNDEPIEM----- 456
373 SVSPPTATSPVLIQTSANR-----SVTFSSKRLSEEEKTPDNLTSQIQPIELKSGR 426
457 -----ANGQPVYIQTLTVEEGEYTGDIYFAK-----GDNV--LYSSLIELSQR 497
427 LVLKRAVLAPSLSODPOLIMEAGTSL-----KTSDDLKATLSLIPLSLDEKS 479
498 IILREQTKLVNLSLTQIGGS-VHMEGCGTLDFAVTPPAANSMAKLTIVHFSLASLKNKC 556
480 VTIHAPNLSIO-----KIFLSNGDENFVENVELLSKFOONNIPLLT 520
557 VTNPTNPPOVSSPAVIGNTAGTATISGPIFEEDL-DETAVDNNOQLGADQ-----T 609
521 LSKRQSL-----HLPDGSLSHFGYQCDWTFSSK-----DSDEGHSLLAMWTF 564
610 IDVLQTLHLGANPPANAPDITLFGNESSKYGYQSWITQWEPEDPANPPONNSYMLKASWTK 669
565 KNYPHREPOSTLVANFLMTVYSDMAVOQMINTIAHGAVLFGTQWSAASNLFFAIDSS 624
670 TGYNPHGERASVLSNLSKMSILDVRSASHAIGASIDCRAYCKRGIWISGISNFTY-HDQD 728
625 CKPIDNMHHNSLGLFCISTHSLDDHSFCLAAGQLLCKSSDSPIYSTET-----TSYIAT 679
729 ALG-OGYRHSISGYSIGANSY-FGSSMFGLAFTETFEGRSDYVVCRRNDHTCVGSVYLT 786
680 VQALALASLM-KISAQACYNESIHLEKTKYRSFEKBEFGSMHSAVSGEVCASIPVINSNG 738

DB 787 RQALSSCLFPGAFAVRASTGFCGNQHKTSY-TRAESNVRWMDNNCVGVGAGCLPIMLAA 845
OY 739 SGLF-SFSIFSKLOQFSGTODGFESSGEIRSFSASSFRLISLPIGTFEKKSOQRTY 797
DB 846 SKYILNELRFVQAEFAVAHEHESFTERGQDAREFKSGHLMNLSPYGVAFDRCSKHPK 905
OY 798 YFELGAVIQDLKADVSGPVLKNAVSDAPANLDSRAVMPRLTNQAL-HRLDTLLN 856
DB 906 YSPMGAVICDPAVRSISGTETTLTSHKETWTTDAFHLARHGVAVGSMVSLGTNIEVGH 965
OY 857 VSCVLRGSHSVSLDGTYYRF 878
DB 966 CKYEVHDASHGCLSIGSKIRF 987

RESULT 16
G71460
probable outer membrane protein G - Chlamydia trachomatis (serotype D, strain UM3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: G71460
R:Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Milche
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A:Reference number: A71570; MUID:9900809
A:Accession: G71460
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1013 <ARN>
A:Cross-references: GB:AE001360; GB:AE001273; NID:g3329342; PIDN:AAC68469.1; PID:g332
A:Experimental source: serotype D, strain UM-3/Cx
C:Genetics:
A:Gene: pmpG

Query Match 16.7%; Score 758; DB 2; Length 1013;
Best Local Similarity 26.2%; Pred. No. 7.8e-40;
Matches 265; Conservative 139; Mismatches 399; Indels 208; Gaps 34;

26 DPLGETA-----LTKRPNHVCTFFEDCTMESLFPALCAHASODPVLVGNISYCFWS 80
53 DPSCGVYFSAGELTLKLNLDNSIALPLSCFGLN-----GSFTVLGNGHSLTFE 101
DB 81 KLHTTDEKAL--FKKQGLSIONFRPLSFTDCSS-----KSSSPS- 119
102 NIRSTNGAALSOSANGLETTFIEGFKELSFSCNCSLAVLPAATNNGSOTPTTSTPSN 161
OY 120 -IHKNGQSLNNNGSMSPCHHABSGGAIADAFSLQHNVLTFAPFENSCKGNGAI 178
162 GTIYSKT-DILLNNEKFSFYSNLVSGDGAIDAKSLTVGISKLVCEONTAODGAC 220
DB 179 QAOT-FSLSRNVSPISFARNRADLNGAICC----- 208
221 QVYISFSAANEAPLAFIANVAGYRGCGIAAYVDGQGVSSSTSTEDPVVFSRNTAVF 280
OY 209 -----SNLSCGVN-----NPLEFTGNSATN-----GCAI 233
DB 281 DGNVAVRGGIGYSYGNVAFNLNGKTLEFLNNVASVYVYIAAOPNTNGOASNTSDNYGDGAI 340
OY 234 CC---ISDLNTEKGSLSLACNOETLPASNAKEKGAITAKHVLRYNCPVSPINNSAK 290
DB 341 FCKNGAQAASNGSVSPFGEGVYFSSNVAAGKGAITAKKLSVANGCGVDFLGNIAN 400
OY 291 ICGAIAIOSGSLSLGEGSVLPFONNSORTSDGLVR-NAIYLEKDAI-----LSL 342
DB 401 DCGATVYGFSGELSLSDADYDIIIFDGNLKRTRAKNADVGVYSSQAISMGSCKRTTL 460
OY 343 EAKNG-DLFFDPPIVOESSKESPLPSLQASVTSPPATASPLVLIQTSANRS--VIFSS 399
DB 461 RAKGHOILFNDEPI-EMANNNQPAOSS-----EPLKINDGEGYGDYIFA- 505
OY 400 ERLSEEEKTPDNLTSQIQPIELKSGRLVLDKRAVLAPSLSODPOLIMEAGTSL--- 456

Oy 487 -LSIQIFLNSDNDENFEYVELLSKEQNNIPILITLSKEQSLLHPD-----GNLSSHF 539
 Db 572 ALTISCTGLVKNKSDCCDNHGHFNKDLQVPILELKAITSNTVTITDPSLGTNGYQDSFY 631
 Oy 540 GYGDWTFESKDSDEGHSLIANWTPKNVPHPEROSTLVANTLNTYSQMAVQSMINTI 599
 Db 632 GYOGTWEFTIDTFT--HTVGNMKKTGYLPHPERLAPLIPKSLMANVYIDLRAVQA--SA 687
 Oy 600 AIG-----GAYLFTQMSAVSNLEYA-HDSSKRPIDNNIHRBLGYLEFGISTHLDHSEFL 654
 Db 688 AGEDVPGKQLST---GTFNFHANITGDAR---SYRMGGGYLINTYTRITTPAALSL 741
 Oy 655 AAGQLLKSSDSFTSTSTETTSYIATVOAOLATSLMKIS-----AOACYNESIHELKTK 707
 Db 742 GFGQLTKSKDYLVGHGHSNVYFATVYSNITKSLFGSSRFSGTSRPTYSNSNEKVATIS 801
 Oy 708 YRSFSKEGFGSMHSAVAVSGEVCASIPVNSGSL-FSSFSIFSKLQSGTODGFEESG 766
 Db 802 YTKLPK-GRCSYNNNCNLGELFGLNLPYTLSSRLNLKQIIPFVKAFVAYATGCIQENRP 860
 Oy 767 EIRSFSSSRNISTPLGITFEKKSKQTRITYYFLGAITDILKRPVSGPVLLKNASW 826
 Db 861 EGFIHGHLINAAVPGVAFRGKSNHNRPFYTIIVAAPVYRINPDCDTLLPLINGATW 920
 Oy 827 DAPMANLDSRAYMFRLTNORALHR-LQTLNVSCVLRCQSHSYSDLGQYTRAF 878
 Db 921 TSIQNLTLSTLLVQASSHTSVNDVLEIFGHGCGDIRTSNQYTLDIGSKLRF 973

 RESULT 12
 C81593
 polymorphic membrane protein G family CP0299 [imported] - Chlamydia pneumoniae (str
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #texi_change 11-May-2000
 C:Accession: C81593
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gilm, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A>Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
 A:Reference number: AB1500; M01D:20150255
 A:Accession: C81593
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-995 <REA>
 A:Cross-references: GB:AE002191; GB:AE002161; NID:g7189216; PIDN:AAF8156.1; PID:g7189222
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: CP0299

```

Query Match      18.0% Score 816.5 Db 2 Length 995
Best Local Similarity 26.4% Pred. No. 1.5e+43
Matches 252; Conservative 158; Mismatches 404; Indels 139; Gaps

QY 37 NNHNHVCFFPEECCTMESLFPAALCAHAS-----QDDPLYVICNSYCVKSLHTIDPREAL 91
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 71 DRRGTLCLIFSGDLYIANDNMIISRTSSCSNRKGAQLIILAKGVSEFLIRSSADGAAT 130
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 92 FR-----EKGLSIONERFLFTDCSSKESSPS-----IIHQ-----NGOLSLRNNGMS 137
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 131 SSVITQNPCLPCLSTSGFSQMFQNCESLISDTSASNVIPHASIIVATTPMLLTNNDSL 190
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 138 FCRNHAEGSGAISADAFSLQHNLYLTFAAFEENSSKGNGAIO-AQTFSLSRNVSPISFAR 196
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 FQYNSAGGAIRGTSTIETKKSLFENGNGSISNGALTGSAIINLINSAPVIFST 250
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 197 NNADLNGAICCS--NLICSGNVPNLPFTGSAINGAICICISDLNISEG----- 245
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 251 NATGIYGAIVLTGSMLTSGNLGVLVNNSSRSGAIVANGVNTVSNNSDLTFOONTA 310
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 246 -----SLSLANQETLFASNAKE 264
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 311 SGNQSLPAPPTPPPTPAVYPLLITGGAIFCTPPAPPTPPPTGVSLTISGNSVTELENTASE 370

```

Oy 265 KCGATVAAKHAMVLRNCPSPFNNNSAKIGCAIAIOSGSLTILAEAGSVLAPFNNSQRPISDQ 324
 Db 371 QGGALYGKKITISDNKSTIFLGNTAGKGAIAIPESEGLSISANOGDILFNKILNISTSGT 430
 Oy 325 GLVRNAIYLKENDAILLSLEARNG-DILFFDPDVOESSKESPLPSLOAVSPSPPTAS 383
 Db 431 P-TRHSIFHCDAKFATLGATQGYTLVFPDPTSDLSAAS-AAAAAAAAAAAAAAAAAAA 473
 Oy 384 PLVIOTSNR-----SVFSSERL-SEEKTPDNLSLOQDPELKSGRVLKDRVLS 436
 Db 474 TVVVAHPKASADGASGVTLVFSGETLTLEAATPANANSTLNOKLELEGTLALNCAATLN 533
 Oy 437 APSLSODQALLINEAGSLTSSD-----LKATLSLPLSHLDPEKSVTHAPN--- 486
 Db 534 VHNFTQDCKSVYINDAGTTLTTLTGANNTDCAITLNLKVLITLDSLDGTGAIAVWVOSTNG 593
 Oy 487 -LSLOKIFLSNGDENFYENVELLSKEONNIPLLTSLSKOESHLPD-----GNLSHF 539
 Db 594 ALTISGLGLGVKNSDCCDNHGMFNKLDQVRILELAKTSNTVTTTDFSLGTNGYQOSPY 653
 Oy 540 GYGDWMTSMKSDSGHSLLANWTPKNTVPIPEROSTIVANTLNTYSDMOAVOSMINTI 599
 Db 654 GYOGTWETITPTT--HVTGNTMKKTGYLPAPERLAPILPMSLANVYIDLAASQA--SA 709
 Oy 600 AHG-----CAVLEFGMGSAVSNLEFA-HDSGSKPDIOMNHRSLGLFGISTHSLDHSFCL 654
 Db 710 ADGEVPEPKQSLST---GITPFNANTGDR---SYNMHGGLINTYTTLITDALS 763
 Oy 655 AAGQLLGKSSDSFISTETTSYATVQAOLATSLMKIS-----AQAQYNESIHLELTK 707
 Db 764 GFGQLFTKSKDYLVCHGSHNVYFATVYNTIKSLFGSSRPFSGGSRVATYSRSEKVKTS 823
 Oy 708 YRSFEKEFGSHVAVSGEVCASIPYVNSGSL-FSSFSIFSKLOGSGTGQDPESSG 766
 Db 824 YTKLPK-GRCSWNNCMWGLEGNLPIPLSRILNLKQIIPFVKAEEVAVATHGIGIQENTP 882
 Oy 767 EIRSFSSASSFRNISLPIGITFEKKSOKTRTYVYFAGATYDCLKRDVESGPAVLLKNNAVSW 826
 Db 883 EGRIGCHGLLVANAPVGVREGRCKNSHNRPDYTLIIAARPVYRHNRPCCDTLLINGATW 942
 Oy 827 DAPMANLDSRAVMPRLTNOALHR-LQTLNVSCVLRGQSHSYSIDLGCTYRF 878
 Db 943 TSGINNLTRSTLIVQASSHTSYNDLEIFGHGCGDIRTSRQYTLIDISKLPF 995

RESULT 13
 C72078
 polymorphic outer membrane protein g family - Chlamydothila pneumoniae (strain CWL0299
 C.Species: Chlamydothila pneumoniae, Chlamydia pneumoniae
 C.Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
 C.Accession: C72078
 R.Katman, S.: Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Ollinger, L.; Grimwood,
 Nature Genet. 21, 385-389, 1999
 A.Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A.Reference number: A72000; MUID:99206606
 A.Accession: C72078
 A.Status: preliminary
 A.Molecule type: DNA
 A.Residues: 1-936 <AR>
 A.Cross-references: GB:AE001627; GB:AE001363; NID:94376721; PIDN:AAD18589.1; PID:943727
 C.Genetics:
 A.Gene: pmp_7

	Query Match	17.1%	Score 773;	DB 2:	Length 936;
	Best Local Similarity	28.0%;	Pred. No. 7.8e-41;		
	Matches 236;	Conservative 143;	Mismatches 376;	Indels 86;	Gaps 22
QY	97 DLSIQNFRLLSTDDCSKSSPS--IIHOKNQGLSRNNNGSHSPFCNNHAEGSGAISAADA	154			
	:: :: :: :: ::				
Db	119 NLEFNDFSLSLIISGCSLLSPGGCCALSVGSLUSTGSQILIFQNFSDDMGIVNTKN	178			

A: Molecule type: DNA
A: Residues: 1-928 <ARN>
A: Cross-references: GB:AE001628; GB:AE001363; NID:94376730; PIDN:ADD18593.1; PID:94376730
A: Experimental source: strain CWL029
C: Genetics:
A: Gene: pmp_11

Query Match 18.6%, Score 845; DB 2; Length 928;
Best Local Similarity 28.8%; Pred. No. 2,26-45;
Matches 247; Conservative 148; Mismatches 345; Indels 118; Gaps 26;

QY 97 DISIONFRFLSPWDOSSKSSPS-IIHOKNQLS-----LRNOSMFCRHHAGCSGA 149
DB 113 NLTEGSEFLSF-----DSSPTVTVTGQGLSSAGGVNLEIRKLVLVAGFEFTADGA 166
QY 150 ISADAFSLQIHWLFTAFEESSKNGCAIOAQTFS-LSRNVSPISFARRADLNGAICC 208
DB 167 IKGASFLLTGTSGDALFESNNSSSTGCAIATTAGARIANNQCYFLSLNASTSGAID 226
QY 209 SNLICSGVNVLPFTGNSA-TNGAICCIISDLNTEKGS--LSLACNOETTFASNAKEK 265
DB 227 EGTSLISNNKFLYFEGNNAKTGGAIC-----NTKASGSPELIISNNKTLIPASVAFES 281
QY 266 GCALIAKHHVLRYNQNPVFINNSAKI-----GCALIAIOSGSGISLAGEGSLVFQNNQRT 321
DB 282 GGAHIAKKAIASSGFTFELRNNVSSATPKGASIDASGELSLAETGNTTFVKNLT 341
QY 322 --SDGGLVNAIYLEKDAIILSLF-ARNODILFPDPIVOESSKESPLSSLOASVTSPT 378
DB 342 TGSTDTPKRNAINISNGKFTFLRAAKNHTTFEYDPTISEGSSDYKINNGSAGALNPY 401
QY 379 PATASPLVYIOTSANKSVIFSSERLSEEE-KTPDNLTSOLOPIELKSGRLVLDRAVISA 437
DB 402 GGT-----ILFSGETLPADELKVDANKLSFTGVSLSGGKLLLOKGVTLFS 448
QY 438 PSLSDPPQALLIMEAGTSIKTSS-DLKLATLSPLHSLDEKSVITHANLSIOKIFLSN 496
DB 449 TSFSQDAGSLDMDSGTLSTTAGSITTNIGLINDSLKOPVSLTAGAS-NKVYISG 507
QY 497 -----SGDENFEYNELLSEKQ-----NNIPLTLSKEGSHLIPGNLSS 537
DB 508 KLNLDIDIEINILYES-HMPSHDOLFSLKITTVAADVDTNVDISL-----IYVPAEDRPS 560
QY 538 HFGYQDMTFWSKSDSEGHNS--LIAMTPKKNVPHPEROSTLVANTLMTYSDMOAYOSM 595
DB 561 EKGFGQGNVNVNTDTATNTKEATATWTKTGFPVSPERKSALVCNTLMOGVFDRISLOL 620
QY 596 INTIAHGCVLEFCTGSAVSNLPIAHDSGKPIDNNHNSLGLVIGTISHSLDHSFCLA 655
DB 621 VEIGATGMEHKOGFWVWSMTN--FLHKTGDNENRKGFRHFSGGVYIGGSAHTEPKDLETPA 678
QY 656 AGOLLGKSSDSFTTSETETSYATV-----QAOLATSLMK----- 690
DB 679 FCHLPAKDKDCIATNNNSTYGGTLEFKHSHLQIPONTYLRIGRAKSESAIEKFPREPL 738
QY 691 -ISAQACYNESIHLEKTKYRFSKREGGSMHSAVAGEVCASIP-IVNSGSGLFSSFSIF 748
DB 739 ALDVQVFSHSDNRMMETHYTSL-PESEGSMSNECIGAGIGLDPFLVSNPHLPKFTFIQ 797
QY 749 SKLQGSSTGQDPEFESSGELKRSFASSPRNISLPICITPEKKSQKRTYTYFLGATIDL 808
DB 798 MVEVENVYVSONSEFESSDGRGFSIGRLNLSIPVAKVVOGIDGSYTYDLSGFVSQV 857
QY 809 KRDEGSPVLLKNAVSMOAPMANLDSRAYMFLTNORALLHRLQTLNATSC-----V 860
DB 858 YRRNPQSTATLVMSPPSKMIRGCNLSROAFLRGSN-----NYVNSNCELFHYAME 910
QY 861 LRQSHSYSLDGLTYR 878
DB 911 LRQSSRNMYNDVQTKLRF 928

Db 811 NFKSGAEARAFDDGDLVNCSTIPVIGILEKISEDEKNNFELSLAYIDVYRKNRSTSL 870
 QY 820 LKNAVSDAPMANLDSRAVYFRLNQRALH-RIOTLLNVSCVLRGOSHSLDGTYYRP 878
 Db 871 MVASASWTSICKNLARQAFPLASAGSHLTLSPHVELSGEAAVELRGSAHYIVNDCGLRYSF 930

RESULT 5

D72078
 polyomphic outer membrane protein g family - Chlamydia pneumoniae (strain CWL029)
 C:Species: Chlamydia pneumoniae
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
 C:Accession: D72078
 R:Kalmun, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
 Nature Genet. 21, 385-389, 1999
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A72000; MUID:99206606
 A:Accession: D72078
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-930 <ARN>
 A:Cross-references: GB:AE001627; GB:AE001363; NID:94376721; PIDN:AND18590.1; PID:9437672
 C:Experimental source: strain CWL029
 C:Genetics:
 A:Gene: pmp_8

Query Match 20.28; Score 915.5; DB 2; Length 930;
 Best Local Similarity 30.84; Pred. No. 7.6e-50;
 Matches 255; Conservative 150; Mismatches 344; Indels 87; Gaps 26;

QY 98 LSIQNFRLFLS----TDCSSKSSPSIIHQKQGLSLRNGNSMFCR---NIAEGSGA 149
 Db 119 LTFGFSNLSPFIAPGTTVASGKSTLS---SAGALNLTQDCTLLFSGQVNSNANNGCA 174
 QY 150 ISADAFSLQHNHYLFTAFEESSKNGGAI-QAOTFSLSRVSPISPARNADLNGAIC 208
 Db 175 ITTKTSLISGNTSITFTSSAKRLGAIYSSAASISGNGOLVFNKKETGGAL- 232
 QY 209 SNLICSGNV--NPFFTGNSATN---GCAICISDINTSEKSLSLAQOETLIFASNS 261
 Db 233 -GFASSSTIQNSSLFFSGNATDAAGKAIYC---EKTEPTLTITSGKSLTFEANS 288
 QY 262 AKERGAIVAKHMYLRNGPVSTINN---SAKIGAIQSGGSLILAGECVLEFON 316
 Db 289 SVTGGGAIACAGLDLSAAGPTLFFSNKCGNTAAGCAIAIADGCSLSLANODITFLG 348
 QY 317 NS-QRTSDQGLVRNATYLEKAILLSLEARNG-DILFPDPIVOSSSKESPLPSLQASV 374
 Db 349 NTLTSTAPSTRNATILGSAKITNLRAGOGSITFFDPIASNTTGASDVL----- 400
 QY 375 TSPPIATASPLVITQTSANRSVIFSSERLS--EEKTPDNLTSOLOQPIELKSGRLVLDRA 433
 Db 401 TINDPDSNPL---DYSGITVFSGEKLSDAKADNFTSLKOPALASGTLALAGCNV 456
 QY 434 VLSAPSLSDQPOALLIMEAGTSLKTSDD-LKLATLSIPLSLDTEKSVTHAP-----N 486
 Db 457 ELVDNGCFOTPGSTLLKQPGTKLADTEALSLTKLVVDLSALEGNKSVSITAGANKTIF 516
 QY 487 LSIQKIFLNSGDNENFENVELLSKEQNNIPLLLSKEQSHLHPDGLLS-----HF 539
 Db 517 LTPSLVLEPQDSSG--NFYESHTINOAFQPLVFAATPAASDIYI-DALLTSPVOTPEPHY 573
 QY 540 GYGQDMFWSMDSDSGHSLIANMTPKNVPPHREOSPLVANTLNNVTSDMAVOSMINTI 599
 Db 574 GYGQHMEETAMDITSTAKSGTWTVTGYNPRERRASVPPSLMASPTDITDQIMTSQ 633
 QY 600 AHGAIVLEFGTWSAVSLFYAHDSGKPIDMHHRSGLYFGISTHSLDHSFCLAGCOL 659
 Db 634 ANSIVQOGLASGTAFFH-KDKSGTN-QAFRRKSYCIIVGSAEDSENIFFVACQL 691
 QY 660 LKSSDSITSTETTSYATATQAO-----LATSLEK-----ISAACYNES 700

Db 692 FGKDKDLFIVENTSHNLTASYLOHRAFLGGLPMPSPGSGITDMKIDPLILNQLSTYT 751
 QY 701 IHELTYYRSPFKSGFSGMWSVAVSGEVCASIP1-VNSGGLFSSFSIFKLOGFSGTQD 759
 Db 752 KNDMDTFRYTSV-PEAQGSWMNNSGALFELGSLALYLKKAEPFPGCYRPLKFAVYSRQ 810
 QY 760 GFEESSGEIRSFSSASRRNLSLPIGITEKKSQKTRYYVFLGAYIODLKRVDSCPVYL 819
 Db 811 NFKSGAEARAFDDGDLVNCSTIPVIGILEKISEDEKNNFELSLAYIDVYRKNRSTSL 870
 QY 820 LKNAVSDAPMANLDSRAVYFRLNQRALH-RIOTLLNVSCVLRGOSHSLDGTYYRP 878
 Db 871 MVASASWTSICKNLARQAFPLASAGSHLTLSPHVELSGEAAVELRGSAHYIVNDCGLRYSF 930

RESULT 6

C81591
 polyomphic membrane protein g family CP0309 [imported] - Chlamydia pneumoniae (s
 C:Species: Chlamydia pneumoniae
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
 C:Accession: C81591
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
 A:Reference number: AB1500; MUID:20150255
 A:Accession: C81591
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1276 <REA>
 A:Cross-references: GB:AE002193; GB:AE002161; NID:97189224; PIDN:AAF30166.1; PID:9718
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: CP0309

Query Match 19.78; Score 894.5; DB 2; Length 1276;
 Best Local Similarity 30.64; Pred. No. 2.7e-48;
 Matches 279; Conservative 144; Mismatches 364; Indels 125; Gaps 29;

QY 68 LTVLGNISYCMVSKL-----HITDPKALEKKEKGLDLSIQNRFSLPDDSSKSSPSIIH 122
 Db 389 LYSKGNSSLTGNTNLLFSGNKATGPSNSANQEG-----CGAIIISLESAS-----VS 437
 QY 123 QKNQGLSLRNGNSMFCRNHAEQSGAISADAFSLQHNHYLFTAFEESSKNGGAIQAT 182
 Db 438 TKKG-LWIEDENENVSLSGNATVAGAIYATKALHGNNTLT-EDGNTAEFGAIYET 495
 QY 183 --FSLSRVSPISPARNADLNGAICCSNLICGANNPLFTNSATN----- 229
 Db 496 EDFTLTGSYGTIVFTPTAKTAGALHTKGNTPFKN-KALVFSGNSAATATTTTDOEGC 554
 QY 230 GGAICCC-ISDINTSEKSLSLAQOETLIFASNAKE----- 264
 Db 555 GGALICNISESDITK-SLTITENESLFTINTKKBGGGIIYAKCVISGSESTINFGNT 613
 QY 265 --KGAIVAKHMYLRNGPVSTINSAKIGAIQSGGSLILAGECVLEFONNSORT 321
 Db 614 AETSQGAIVSKNLSITANGPVSTINSGCGGAIYIADSGSLAEALDIDITFGCN--RA 671
 QY 322 SDQGLVRNATYLEKDALLSLEARNG-DILFPDPIVOSSS-----KESPLPSLQASV 375
 Db 672 TEGTSTPNSIHLGAGKITKLAAPGHTIYFYDITMEAPASGCTIEBVLNPPVKAIVP 731
 QY 376 SPT-----PATASPLVITQTSANRSVIFSSERL-EEKTPDNLTSOLOQPIELKSGRL 427
 Db 732 PPQKNGPIASVPVPAAPANPNNGITVIFSSGKLPSQDASIPANTTTILNOKILAGGAV 791
 QY 428 VLKDAVLSAPSLSDQPOALLIMEAGTSLKTS-----DLKATLSIPLSLDTEKSVT 481
 Db 792 VLKEGATLQVYSFTQGPDSVTFVMDAGTTLTETTTNNNDGSLDLNLSVNDALDQKMIT 851
 QY 482 I-----HAPNLSIQKIFLNSGDNENFENVELLSKEQNNIPLLTSLKEQSHLHPDGN--- 534

Qy	421	ELKSGRLVKRRAVLKRAVLSLPDPOALLIMVAGTSLKTSDDLKATLSTPLHSLDTEKSV	480
Db	421	ELKSGRLVKRRAVLKRAVLSLPDPOALLIMVAGTSLKTSDDLKATLSTPLHSLDTEKSV	480
Qy	481	THAPNLSTIOKIFLPSNSGDENFENVVELLSKEONNIPILTJTSKROSHLPLDGNLSSHFG	540
Db	481	THAPNLSTIOKIFLPSNSGDENFENVVELLSKEONNIPILTJTSKROSHLPLDGNLSSHFG	540
Qy	541	YOGDWTFMSKSDDEGHSLLANWTPKNTVPHPEROSTLVANTLWNTYSDMQAVQSMINT1A	600
Db	541	YOGDWTFMSKSDDEGHSLLANWTPKNTVPHPEROSTLVANTLWNTYSDMQAVQSMINT1A	600
Qy	601	HGGAVFCTGWSAAYNLPEYAHDDSSGKPIDNMHHSLSLGLFSTHSLDDHSPCLAAQGLL	660
Db	601	HGGAVFCTGWSAAYNLPEYAHDDSSGKPIDNMHHSLSLGLFSTHSLDDHSPCLAAQGLL	660
Qy	661	GKSSDSFTTSTETTSYIATVQAOALATSLMKISAQACYNES1HELKTKYRSPSKGFGSMH	720
Db	661	GKSSDSFTTSTETTSYIATVQAOALATPLMKISAQACYNES1HELKTKYRSPSKGFGSMH	720
Qy	721	SVAVSGEVCASITPVVNSGGLFSSPS1FSKIQGFSQNDCEEBESGCELRSPSASFPRLS	780
Db	721	SVAVSGEVCASITPVVNSGGLFSSPS1FSKIQGFSQNDCEEBESGCELRSPSASFPRLS	780
Qy	781	LP1G1TFEKKSKOKTRTYYYFLGAYIODLKRDESGPVLLKNAYSWDAPMANLDSRAVME	840
Db	781	LPMG1TFEKKSKOKTRNRYYYFLGAYIODLKRDESGPVLLKNAYSWDAPMANLDSRAVME	840
Qy	841	RLTNORALHRLQTLNLNVCVLRQGSHTSLDGLTTPRR	878
Db	841	RLTNORALHRLQTLNLNVCVLRQGSHTSLDGLTTPRR	878

RESULT 2

F81721

polymorphic membrane protein G family TC0267 [imported] - Chlamydia muridarum (strain Nijmegen)

C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #texl_change 11-May-2000

C:Accession: F81721

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heldberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, S.L.

N:NCIC Acid. Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.

A:Reference number: A81500; MUID:20150255

A:Accession: F81721

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-867 <TE>

A:Cross-references: GB:AE002294; GB:AE002160; NID:g7190305; PIDN:AAF39136.1; PID:g7190305

A:Experimental source: Strain N19g (MoPn)

C:Genetics:

C:Gene: TC0267

	Query Match	78.2%	Score 3543.5	DB 2	Length 867
	Best Local Similarity	77.8%	Pred. No. 5.7e-216		
	Matches 684	Conservative 76	Mismatches 106	Indels 13	Gaps 2
QY	1	MRPDHMFCCCAALISLSTAVLFCQDPLGELTALTTKNHNHVVCTFEEDCTMESLFPALCA	60		
DB	1	MRPDHVNLCCLCATILSTPTALIFGQDALDKSLITKKNNSIVCFLEDCCTMENSPLALS	60		
QY	61	MASQDDPLVLCNLSICWYSCSKLHITDPKEALFKKEKGDISIQNFRLLSTTDCSSK-ESSPS	119		
DB	61	MARQDDPLVLIQNTNMHWVSNLHPSTNEERFLKEKGDISIQDFRLSTPTDCSSSTEPS	120		
QY	120	IIHOKNGQLSLRNNGSMFSCRHHAEGSGAISADAFSLQHNHVLFTAFEPENSSKNGGAIQ	179		
DB	121	IIYHKNQGLFLRNNGNMMSFYRNHSGSGCALSTDBLFLQHNHVLFTNEPENSAAKNGGAIQ	180		
QY	180	AQFSLSRNVSPISFARNRADLNGCAICCSNLIICSGNVNPLFTGNSATNGAICCIISDL	239		
DB	181	AQTLSLSRWVSLSSSRNANLNGAICCONICSGNVNPLFTTNSLNGAICCIINQ	240		

```

QY 240 NTSEKGSLSLACNOETPLFASNSAKEKGAIYAKHMVLRKYNPVSFINSKATIGAINIOS 299
Db 241 NLSEKGSLSLAVNOETPLFSGNSAKEKGAIYTKHVLRHNPVSFVNNASAKLGAIAIOS 300
QY 300 GGSLSILAGESSVLFQONNSORTSDGGLVRNAYIEKDALISLEARNBDILFFPPIAOES 359
Db 301 GGSLSIINGGSSVLFQONNSCHFSDQGYRNAYIYEKANALLSLEARNBDILFFPPIAOEV 360
QY 360 SSKESPLPSSLOASVTSFPTWASPHVLOTISANSVLFSSERLSSEEEKTPDNLTSOLOQ 419
Db 361 VSPSESTTSAL-----TPLRIOTNRNAVIFSSENLSKEEKEATEMLTSKIQP 408
QY 420 TELSGRVLVDRAYLASPLSOPBOALLIMEDATSLKTSSDLATLSTPLSHLDEKS 479
Db 409 IELSGCVCVLDRVILSPLSLOAPQALVMDGVSTLSSDLKLTTLSTPLSHLDEKS 468
QY 480 VTHAPNLISOKIFLNSGDNFENVELLSKEONNIPILTLSEKQSHLHPDGNLSHF 539
Db 469 VSIQSPFLISOKIFLNSGDNFENVELLSKDKDIDPLSLPGLPDPDGNLSHF 528
QY 540 GYQCDWTSKSDSDSGHLLANWPKNVPRPEROSTLVANTLWNTYSDMOAVOSMINTI 599
Db 529 GYQGMNNSWOTSDRELVLANWNTANSYIPRPEROSALVANTLWNTYSDMOAVOSMINTI 588
QY 600 AHHGAYLEFGTMSAVSNLFYAHSDSGKPIDMHHNSLYLGLGISTHSDHDSFCLAGOL 659
Db 589 AOGGAYLEFGTMSAVSNLFYSHGNSGKSTDMWKKHRSLOLFCISTHSDHDSFCLAGOL 648
QY 660 LKSSDSFTTSTETTSYATVQAOALATSLMKISAOCYNESIHLELTKYRSKREGQSW 719
Db 649 FGKSSDSFVTADTTSYATAIOTQATSLIKISAOCYNESIHLELTKYRSKREGQSW 708
QY 720 HSWAASGEVCAISIPVNSGSGLFSSSIFSKLQGGSGQODFEESSGCIIRFSKSSPFINI 779
Db 709 HSWAASGEIGASIPVNSGSGLFSSSIFSKLQGGSGQODFEEBSRGCAFAAUFSSFINI 768
QY 780 SLPLITTEKKSOKRTTYVFLGAYIQDLKRDDVESGPVLLKKNVSWDAPANLDSRAYM 839
Db 769 SLPVGIAIEKKKSQKTRNYVHFLGAYIQDLKRCVBSGPVLLKKNSTWDAAPANLDSRAYM 828
QY 840 FRLTNORALHRLQTLNASCVLRGOSHSYSLDLGTTTYF 878
Db 839 FRLTNORALHRLQTLNNSYMLRGOSYSYSLDLGTTTYF 867

```

RESULT 3
B72077
polymorphic membrane protein G family CP0306 [imported] - Chlamydothrix pneumoniae (s
C.Species: Chlamydothrix pneumoniae, Chlamydothrix pneumoniae
C.Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C.Accession: B72077, B81592
R.Katman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
Nature genet. 21, 385-389, 1999
A.Title: Comparative genomes of Chlamydothrix pneumoniae and C. trachomatis.
A.Reference number: A72000; MUID:99206606
A.Accession: B72077
A.Molecule type: DNA
A.Residues: 1-928 <ARN>
A.Cross-references: GB:AE001628; GB:AE001363; NID:94376730; PIDN:AAD18591.1; PID:94377
A.Experimental source: strain CML029
R.Creed, T.D.; Brumham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg
Nucleic Acids Res. 28, 1397-1406, 2000
A.Title: Genome sequences of Chlamydothrix trachomatis Mopn and Chlamydothrix pneumoniae AR399
A.Accession: B81592
A.Molecule type: DNA
A.Residues: 1-928 <REA>
A.Cross-references: GB:AE002192; GB:AE002161; NID:97189226; PIDN:AAF38163.1; PID:9718
A.Experimental source: strain AR39, HL cells
C.Genetics:
A:Gene: pmp_-9; CP0306